

DOCUMENT RESUME

ED 250 652

CS 007 722

AUTHOR Engelmann, Siegfried; Steely, Donald
TITLE Implementation of Basal Reading in Grades 4-6. Final Report.
INSTITUTION Engelmann-Becker Corp., Eugene, OR.
PUB DATE 80
NOTE 80p.; For a related document, see CS 007 723.
PUB TYPE Reports - Research/Technical (143) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Achievement; *Basal Reading; Classroom Research; Comparative Analysis; Content Analysis; Intermediate Grades; Program Descriptions; Questionnaires; Reading Achievement; *Reading Comprehension; *Reading Instruction; *Reading Research; Teacher Attitudes; *Teacher Behavior; *Teacher Effectiveness; Teaching Methods

ABSTRACT

The results of a three-part study, conducted to present an indepth look at reading comprehension instruction in the middle grades, are described in this report. Four major areas are covered: teacher's verbal reports, program specifications, teaching behavior, and student mastery. The report first describes the research design, or how a profile of reading instruction was established for grades four through six. The procedures of obtaining data are described: audio interviews, video taping, program analysis, and student performance. The results are presented and discussed and the following types of comparisons are made: (1) some teacher verbal reports are compared to actual program specifications; (2) some teacher behaviors are compared to both program specifications and verbal reports; and (3) student performance is compared to program specifications, teacher behavior, and teacher verbal reports. A summary of the study is offered and the results of administrator interviews are presented. The questionnaire used is described and data from questionnaire responses are compared with those gathered from teacher observation. (HOD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

FINAL REPORT

Implementation of Basal Reading
in Grades 4-6

Contractor: Engelmann-Becker Corporation

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Siegfried Engelmann

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

ED250652

05007722

Project Directors: Siegfried Engelmann and Donald Steely

Assistants: Fran Goode
Joe Kuhn
Sarah Norris
Randy Sprick
Paul Williams

Coordinator: Donald Steely

**Experimental
Design:** Wesley Becker
Donald Steely
Siegfried Engelmann

Interviewers: Jean Osborn
Adrianne Allen

Table of Contents

	<u>Page</u>
I. Design	3
II. Procedures Section	8
a. Audio interview	10
b. Video taping	11
c. Program analysis	13
d. Student performance	16
III. Results and Discussion	18
a. Program analysis results for main idea	18
b. Teacher verbal reports	
1. audio interview results	26
2. differences between grade levels	38
3. teacher behavior compared with program analysis data	41
c. Teacher behavior results	
1. teacher behavior and lesson analysis results	45
2. teacher behavior compared with program analysis results	52
3. teacher behavior compared with teacher reports	55
d. Student performance	
1. student performance results	58
2. student performance results compared with program analysis results	61
3. student performance compared with teacher reports	62
4. student performance compared with teacher behavior results	65
IV. Summary	67
V. Administrator Interview	
a. Administrator interview results	68
b. Administrator interview results compared with teacher behavior	72
c. Administrator interview results compared with teacher reports	73
VI. Questionnaire	
a. Procedures	78
b. Results	78
c. Questionnaire results compared with observed teacher	86
d. Implications	91

INTRODUCTION

The purpose of this research is to provide relevant information about what kind of reading instruction typically takes place in grades 4 through 6. The goal, hopefully, is to provide a knowledge base that may serve as a foundation for a marketing strategy.

The research proceeds in three phases. The first phase involved an intensive analysis of how teachers actually teach. This analysis involves examining not only teaching behavior but also the specifications of the programs they use and the student outcomes that result from the teachings. The first phase provides a factual foundation for determining subsequent relationships. Phase two deals with these relationships. Traditionally, teachers' verbal reports have been taken as valid indicators of teacher behavior and of student outcomes. How valid are these verbal reports? Specifically, how much do teachers actually know about the program they use? How accurate are they in reporting on what they actually do when they teach, how they correct, when they reinforce, and how they change procedures specified by the program?

Phase three considers relationships that are even less direct. Phase three deals with two primary questions: a) To what extent does a larger population of teachers produce the same pattern of verbal responses produced by the smaller group of intensively observed teachers? b) What kinds of responses may serve as key predictors or indicators of particular behavior patterns?

The methodological problems associated with this project are astronomical because most have never been dealt with and therefore require invention, validation, reliability, specification of coding procedures, and procedures that permit some sort of computer analysis and comparison with other items. For instance, what constitutes a question that is irrelevant to a topic that is being taught? How do we "score" the spurious prompts presented by

the teacher? How do we provide a score for the extent to which the teacher deviates from the program specifications? How do we judge whether a program contains distractors that may reduce the probability of correct responses?

There are similar questions associated with all facets of the analysis, even what might seem to be the most straightforward. For example, how do we determine how long it has been since a particular skill had been presented to the students? The lessons of the typical basal are not labelled, so we don't have a ready-made measure for comparing the number of school days with the number of specified lessons. Instead, we must invent convention for determining the number of lessons.

One point must be made clear at the beginning of this report. The designers are quite aware that some of the conventions designed for this research serve as less than perfect indicators of the different criteria examined; however, the designers tried to deal with the issues that are relevant, not merely with the issues that are easy to measure or code. As a result, the current creation of over 60 categories or criteria for evaluating teaching or programs that have never been reported in the literature on instructional design. Certainly some of these would benefit from refinement; however, reliability tests have been conducted for all categories and all are 80% or above--even for criteria that are difficult to apply. The average of all reliability scores is over 90%.

DESIGN

To provide a profile of reading instruction in grades 4 through 6, we must take into account an almost unwieldy amount of information. The primary questions that we must answer include:

- How much do the teachers know about the specific content (or skills) they are attempting to teach?
- How reliable are teachers' reports about what they do to teach particular skills?
- How valid or effective are the procedures teachers typically use?
- How closely do teachers follow the instructional material they use?
- What are the actual student outcomes associated with the teaching of specific skills?

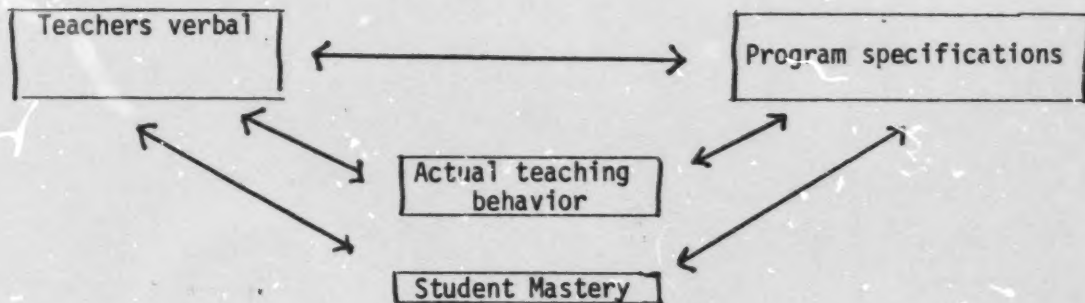
In addition, we must deal with predictions:

- To what extent are student outcomes predicted by the material that is used or by what the teacher actually does to teach the skill?
- To what extent do the teachers' verbal responses predict (1) what the teacher actually does to teach the skill?
- To what extent does the teachers' responses to a written questionnaire predict their actual teaching behavior and predict student outcomes?

The present study is not capable of providing definitive answers to all questions; however, Figure 1 shows the primary predictions dealt with. Four major areas (represented by boxes) are involved: (1) teachers' verbal reports; (2) program specifications; (3) teaching behavior, and (4) student mastery. The direction of the arrow indicates the direction of possible predictions. For example, the arrow from program specifications to student mastery indicates that by examining the content of the program, we might be able to make predictions about the student outcomes. Similarly, the arrow from teachers' verbal reports to actual teaching behavior indicates that it may be possible to make predictions of teaching behavior from teachers'

verbal reports about how they teach. (This is not to say that what they say they do will correspond to what they actually do, but that some of the things they report are predictive or correlated with specific teaching behaviors.)

Figure 1



The length of the arrows provide a rough indication of how strong the prediction might be. The program specification should be capable of "predicting" actual teacher behavior somewhat better than they predict student mastery, simply because they are further removed from student mastery.

Sources of data

For each cell shown in Figure 1, we tried to secure the most direct information.

Teachers' verbal reports were secured through two one-on-one audio-taped interviews. One interview occurred before the video-taped sessions of the teacher teaching. The other occurred after the taping had been completed.

The actual teaching behavior of teachers was assessed through video records of 17 teachers, each teaching two different reading lessons.

The student mastery associated with their teaching was assessed through student material designed to test what the teachers taught during one (or more) of the video-taped sessions.

Evaluation of program specification was based on an analysis of the lessons taught to the children (number of examples, specific instructional

3. A commitment was made to participate in the study without engaging in school board meetings.

Bridgeport, Connecticut, has a school population of 22,000 students. All Bridgeport teachers were from the same school, in which 25 percent of students were White, 36 percent were Black, and 34 percent were Spanish. Seventy percent of the school population qualified for the free lunch program.

In Dayton, two schools were used. One school had 90 percent low income and was one of the top 10 Title I priority schools. The population was Black and Appalachian. The other school in Dayton was a carry-over Title I school that was expected to lose eligibility. Students in that school were placed by parent choice and ranged from ADC children to children of the Board of Education members.

The students in the Sacramento school studied averaged middle-to-low income. Although the school was not a Title I priority school, 14 percent of the students' parents qualified for Welfare and 33 percent of the students qualified for the free lunch program.

The basal programs used in these communities are: Houghton Mifflin, Discoveries (Dayton); Holt, Basic Systems (grade 4); and Scott Foresman, Open Highways (grades 5 and 6, Bridgeport); and Ginn, 720 (Sacramento).

The Teachers

Teachers who participated in the study were volunteers who had used basal programs at least one year. All 17 teachers who volunteered were used. Each was compensated with a \$150 voucher for SRA material. Following is a summary of the grade level, sex, and experience data for the teachers. As the summary indicates, the selection was slightly weighted toward fourth grade (With seven of the teachers teaching fourth grade).

<u>Teacher</u>	<u>Site</u>	<u>Grade</u>	<u>Sex</u>	<u>Number of Years Teacher</u>
1	Bridgeport	4	F	3
2	Bridgeport	4	F	1
3	Bridgeport	5	M	5
4	Bridgeport	5	M	3
5	Bridgeport	6	F	22
6	Bridgeport	6	F	4
7	Dayton	4	M	4
8	Dayton	5	F	14
9	Dayton	5	F	8
10	Dayton	6	M	4
11	Sacramento	4	M	1
12	Sacramento	4	F	8
13	Sacramento	4-5	M	3
14	Sacramento	4-5	F	7
15	Sacramento	5-6	M	6
16	Sacramento	F	13	
17	Sacramento	6	F	15

Audio Interviews

The first audio interview was conducted during November. Each teacher was interviewed for a period of 40 minutes. The interviewer presented nine questions that dealt with the teaching of main idea and cause and effect. The same series of questions was asked about one topic of the teacher's choice.

The second audio interview, conducted after the video taping, occurred in March and April. Interviewers asked each teacher twenty questions, most of which referred to the second lesson that had been video taped.

The questions presented in the first interview dealt with:

1. Specific facts about the instructional program the teacher used (such

as how many lessons teach cause and effect, and what kind of workbook exercises accompany the lessons)

2. Judgments about the presentation as prescribed by the program (whether, in the judgments of the teacher, the topic of cause/effect is actually taught or simply introduced)
3. Information about how closely the teacher follows the prescribed presentation (and what kind of changes the teacher makes)
4. Facts about student mastery (what percentage of the students learn the skill)
5. Ability to construct positive and negative examples of the topic (how one would teach that a black cat crossing one's path does not cause bad luck)
6. Information about how the teacher corrects specific mistakes

The second teacher interview asked about:

1. The percentage of students who mastered the skill presented in one of the taped lessons
2. The nature of additional practice needed for those who did not learn
3. The procedures used by the teacher (such as the use of individual rather than group turns)
4. Judgments about the need for additional student feedback
5. Rationale for particular changes the teacher made in the lesson
6. Judgments about correction procedures
7. Diagnostic approaches used by the teachers (assumptions about other members of the group when a high-performing student misses an item)
8. Opinions about DISTAR students

A summary of the questions presented during the first and second audio interviews appears in Appendix II and Appendix III. These appendices indicate how the interviews are scored and provide sample responses for the various scoring categories.

Video-Taping

Each teacher was taped twice during the period November 15 through March 17. The first tape recorded the lesson the teacher had scheduled for that day. The second taping dealt with a topic that demonstrated instruction in a specific skill area (as opposed to general discussion or unidentifiable skills). Taped segments varied from 15 to 35 minutes. Below is a summary of the topics taped for each teacher.

<u>Teacher</u>	<u>Topic A</u>	<u>Topic B</u>
1	Story types	Main idea
2	Main idea	Map skills
3	Story comprehension	Main idea
4	Main idea	Story comprehension
5	Vocabulary	Context clues
6	Story comprehension	Main idea
7	Parts of a book	Parts of a book
8	Parts of a book	Main idea
9	Main idea	Map skills
10	Map skills	Main idea
11	Communications lab	Main idea
12	Vocabulary	Main idea
13	Vocabulary	Keywords
14	Dictionary skills	Inferences
15	Verbs	Context clues
16	Metaphor/simile	Main idea
17	Main idea	Cross references

As the summary indicates, eight usable tapes on main idea and two usable tapes on context clues were available for analysis on the second taping.

Program Analysis

The total teaching and testing provided for main idea was analyzed in grades 4 through 6 for Houghton Mifflin and Ginn 720. Because Bridgeport used two different basals, the fourth grade level of Holt and the fifth and sixth of Scott Foresman (Open Highways and Cavalcades) were analyzed.

In addition to the main idea analysis, two lessons on context clues that had been video taped were analyzed.

The analysis of main idea dealt with every example presented in the appropriate levels of the reading series. The analysis dealt with the nature of the examples (how many, how frequent, what type,) and whether the examples were taught. (If no teacher-directed discussion or questions were provided for a particular type of task within 60 teaching days, the task was considered not taught.) The analysis also considered whether the instructions were ambiguous or misleading, whether the information presented to the students was false, whether the tasks presented were relevant to main idea and whether there were interpretations other than the intended one conveyed by the set of examples for main idea. Finally, the analysis considered whether responses were prompted or led by the teacher, whether the response forms were consistent from example to example, and whether the responses were relatively strong.

On page 19 is a list of the 24 variables that were used in the program analysis. The conventions for defining, operationalizing and scoring most of them were designed for the present project. As a group, the variables provide a rather thorough profile of what the program does and where the program will probably fail. A complete description of the variables and how they were applied appears in Appendix V.

Teacher Behavior Analysis

The video tapes were analyzed to determine (1) precisely what the teacher did when presenting the lesson and (2) the extent to which the teacher's presentation deviated from that specified by the program.

All the teacher's responses during a specified period were classified and counted. The total number of tasks presented provided a basis for determining the teacher's rate of presentation. The types of tasks formed the basis for determining the percentage of correct student responses, the percentage of student errors, the percentage of errors corrected by the teacher, the percentage of tasks presented to individual students, and the percentage of tasks that were reinforced.

Three measures determined how closely the teacher followed the program specifications: (1) the percentage increase of questions asked (based on the number specified by the program); (2) the degree to which the teacher followed the sequential order of activities specified by the program; (3) the extent to which the teacher questions were of the same level of "fidelity" as the program-specified questions (based on ambiguity, relevance, prompts, misleading qualities, and spurious cues for responses).

The list below presents the categories that were scored for each video-taped presentation. Appendix IV provides a description of the categories and illustrations of their application.

Total percentage of questions with errors.

Total percentage of questions that were group tests.

Total percentage of questions that were individual.

Total percentage of errors that were corrected.

Total percentage of errors that were corrected and retested.

Total percentage of tasks that were models.

Total percentage of tasks that were leads.

Total percentage of tasks that were models or leads.

Total percentage of responses that were given general praise.

Total percentage of responses that were given behavior-specific praise.

Total percentage of responses that were praised.

Total percentage of responses that were given negative feedback.

Rate of responses.

Total number of questions.

Percentage of questions that were ambiguous or misleading.

Percentage of questions that were relevant to the concept.

Percentage of probability of correct interpretation.

Percentage of strength of student responses.

Percentage of prompted responses.

Student Performance Analysis

As a follow-up to the second video-taping (January 15--March 17, 1979) test worksheets were presented to each class. The test worksheets were designed to be fair tests of what had been presented during the second taped lesson. They were given to the teacher within two days after the lesson. Students worked the items on the test worksheets independently. The worksheets were scored by project personnel.

The objectives of the worksheets were to assess: (1) the degree to which students could master the skills assumed to be taught through the lesson; (2) the extent to which students could perform on preskills clearly assumed by the lesson.

The design of the worksheets was based solely on the instructions, examples, question forms, and concepts presented (or assumed) by the lesson. The guideline applied was that if something is tested by an instructional program, and if students are expected to perform independently, the program assumes that the students have been taught skills necessary to perform independently. Until an independent test is provided by the program, however, the skill is still assumed to be "untaught" and should not be tested.

For example, the second lesson taught by teachers 5 and 15 dealt with context clues. Students were expected to read sentences and choose the

word that served as a synonym for one word in the sentence. If students were not sure of the meaning of the word in the sentence, they were supposed to look up the word in the glossary. Since the presentation assumed that students know when to look up a word in the glossary, the test worksheet designed for lessons included items that test the students' knowledge of when they should look up a word. (The sheet also tested the students' ability to choose the correct definitions.) Test items were either as identical as possible to those in the program--same length, same vocabulary, same response forms--or were the simplest form of the skill that could be tested (no tricks).

Samples of the various worksheets created for the study and a description of the criteria for designing worksheet items appears in Appendix X.

RESULTS AND DISCUSSIONS

The present study generated an enormous amount of data, for the four major categories--teachers' verbal reports, program specifications, teachers' behavior, and student outcomes. In all, over 200 classifications were analyzed and scored.

The purpose of this section is to provide a manageable presentation of the results. Therefore, not all data relating to the four major categories will be presented. We will discuss one category at a time, starting with the program specification. The discussion will first present mean data for all programs, and then present profile data for each of the programs.

Each of the remaining categories will be treated in a similar manner, but with one addition. After the second category (teacher verbal reports) has been discussed, we will compare some verbal reports to actual program specifications. After the third category (teacher behavior) has been discussed, we will compare some teacher behavior to both program specifications and to verbal reports. Following the discussions of student performance, we will compare it to program specifications, teacher behavior, and teacher reports. Supplement 1 provides data not presented in this section.

PROGRAM ANALYSIS DATA FOR MAIN IDEA

Means for All Programs

The first column of Table I presents the means for the 24 criteria used to analyze the various programs' treatment of main idea.

The second column of Table I presents judgments of the "ideal" that should be achieved by an instructional program. As the number of 100% and 0% entries in the ideal column suggests, the evaluation is heavily weighted with criteria that are judged to be either perfectly necessary or valueless. Note that there could be reasonable exceptions to any 100% or 0% criteria in the ideal column; however, the general rule for these criteria is that they should not deviate greatly from 100% or 0%.

Other "ideals" are expressed as a range. These ranges are arbitrary and could be extended in specific situations. For four criteria, a question mark appears in the ideal column because of the great range of possible variation. The purpose of all designations in the ideal column is to provide a point of reference for evaluating the programs, not to express a precise, inviolable value.

The programs are far from the ideal with respect to the percentage of examples taught in the teacher presentation and the percentage that are either not taught or are ambiguous. An example is considered "not taught" if the question of the type asked about the example had not been presented in fifty teaching days. Ideally, 100% of the examples would be taught. The fact that only 14% of the examples in the programs analyzed were "taught" implied that the program sequence does not provide for realistic practice and review for the different item types.

Ambiguous and not-taught examples comprise 88% of all examples. Ideally, they would comprise 0%. The mean for examples that are either misleading or for which the wrong answer is given is 12%, compared to an ideal of 0%. Variation of teacher presentation wording and variation in items depend

Table 1
Program Analysis Results Across Programs

	Means Across Program	Ideal
1. Percentage of examples taught	14	100
2. Percentage of questions ambiguous and not taught	88	0
3. Percentage of questions misleading and wrong	12	0
4. Percentage of minimum discrimination not taught	5	0
5. Percentage of variation in teacher presentation wording	14	0-15
6. Percentage of variation in student workbook wording	44	0-15
7. Percentage of variation in items, teacher presentation	10	0-50
8. Percentage of variation in items, student workbook	18	0-50
9. Percentage of questions relevant to concept, teacher presentation	62	100
10. Percentage of questions relevant to concept, student workbook	75	100
11. Percentage of probability of correct interpretation	27	100
12. Percentage of response variation	13	0-50
13. Percentage of visual distraction, student workbook	25	0
14. Percentage of academic distraction, teacher presentation, student wkbk.	31	0
15. Percentage of strength of teacher presentation responses	89	100
16. Percentage of strength of student workbook responses	72	100
17. Percentage of prompted, teacher presentation	24	0
18. Percentage of prompted, student workbook	49	0
19. Days since two examples were presented	62	?
20. Total number of examples in program	66	?
21. Number of student examples on same day as teacher material	9	?
22. Percent of student examples on same day as teacher material	14	?
23. Total number of lessons	22	50-80
24. Percentage of examples for which correction in specified	0	100

greatly on instruction-design strategies. It would certainly be possible to design programs that have great variety in the wording of tasks that present examples of main idea. The systematic control of wording and of response variation would be reflected in a program's total number of lessons, days since two examples of a particular type have been introduced, and other criteria that deal with the control of the communication between instructional material and students. The lack of control in these areas, manifested by the programs analyzed, reinforces the notion that the variations in the wording of student worksheet items (44%) represents a serious instructional flaw.

As criteria 9, 10, and 11 indicate, the instruction provided by the programs is neither well-focused nor carefully engineered. Approximately one-third of all questions purportedly dealing with main idea are not relevant to main idea. Furthermore, the lack of control of features of the examples renders a very strong probability that the learner could learn to attend to some accidental feature of the examples (because all the examples have the same accidental feature and therefore the feature becomes a possible basis for identifying the examples). Perhaps the strongest tenet of careful instruction is that the communication with students must be consistent with one and only one interpretation. The programs analyzed presented an average of about four possible interpretations (or a 27% probability of achieving the correct interpretation).

The visual distractions and academic distractions should ideally be 0%. In the programs analyzed, they were 25% and 31%. These distractions serve no positive function with respect to the mastery of main idea; however, they imply to the learner that they are somehow relevant - an unfortunate implication. (The learner is not in a position to judge the relevance of a distraction until considering it.) The strength of responses obtained by the programs appears higher than it actually is. The 72% strength for student workbook

responses suggests that the majority of workbook tasks for main idea are multiple-choice items. Students rarely deal with "open ended" questions, in which they formulate the main idea. This level of performance, however, is the only one that is reasonable for main idea. In what kind of situations would one "summarize" a passage and be provided with three or four choices? Since the programs rely heavily on multiple-choice items, they fail to teach the central skill of "summarizing" or making up the main idea. Instead, they teach the skill of choosing the main idea - a rather limited skill. (And one that would not be expected to transfer to the production of main idea.)

Criteria 19-23 deal with program details for which there is great possible variation. The number of days since two or more examples of a particular item type have been presented is excessive, suggesting that the programs do not review or use what is taught. This situation implies to students that what is taught is not very useful. The presentation of only 66 examples over a three-year period is critically low. So is the total number of lessons (22).

The predictions based on the performance of the programs analyzed should be that they are perfectly incapable of teaching the concept of main idea. They are designed with no apparent understanding of control of task wording, control of item features, control of integration and review of what is taught, and control of the response forms called for in the teacher and student material. In summary, they lack "empathy." They seem to be designed with no consideration that the learner does not understand the concepts being taught and that the learner's task will be far more difficult if the presentation of examples is consistent with more than one interpretation, if irrelevant tasks are presented, if inadequate amounts of practice are provided, and if the transition from one form of activity to the next is not made gradual and manageable.

Within Program Variations

As Table 2 shows, there is substantial variation in most criteria across the different programs. Following is a brief profile summary of each program.

Houghton Mifflin's performance is weak on most criteria. It presents examples of main idea (or outlining of relevant details for main idea) less frequently than any other programs (158 teaching days between lessons on which two or more examples appear). It also has the fewest number of lessons (5). It has the highest percentage of questions that are either misleading or for which the specified answer is wrong (26% of all teacher-presentation items). Finally, it has the lowest probability that students will achieve the correct interpretation from the presentation (21%).

On the positive side, Houghton Mifflin prompts the lowest percentage of student workbook examples (26%) and provides the strongest student workbook responses (87%).

Ginn's positive features are: (1) the relatively short period of time between presentation of two or more examples of the concept (24 days); (2) the total number of teacher-presented examples (139); (3) the total number of lessons (35); and (4) the fact that 100 percent of the teachers' questions are relevant. The program's most serious weaknesses are the low percentages of examples taught and the high percentage of questions that are ambiguous or for which the specified answer is wrong. One hundred percent of the questions are either ambiguous or not taught, suggesting that the learner is not given direction about the correct strategy for handling the concept. The program's assumption seems to be that if the learner is simply questioned about main idea frequently enough, the learner will master the concept.

Scott Foresman is stronger than the other programs with respect to task relevance. One hundred percent of the student workbook questions are relevant to the concept. There is almost no visual distraction (5%) and the

Table 2
Program Analysis Results by Programs

	Houghton Mifflin	Ginn	Scott Foresman	Holt
1. Percentage of examples taught	46	1	0	13
2. Percentage of questions ambiguous and not taught	67	100	100	73
3. Percentage of questions misleading and wrong	26	11	0	8
4. Percentage of minimum discrimination not taught	7	0	7	16
5. Percentage of variation in teacher presentation wording	28	0	13	36
6. Percentage of variation in student workbook wording	50	31	60	54
7. Percentage of variation in items, teacher presentation	36	2	0	4
8. Percentage of variation in items, student workbook	38	16	4	3
9. Percentage of questions relevant to concept, teacher presentation	38	100	87	87
10. Percentage of questions relevant to concept, student workbook	57	82	100	48
11. Percentage of probability of correct interpretation	21	31	26	24
12. Percentage of response variation	30	10	0	8
13. Percentage of visual distraction, student workbook	39	14	5	23
14. Percentage of academic distraction, teacher presentation, student workbook	41	25	22	47
15. Percentage of strength of teacher presentation responses	85	86	100	88
16. Percentage of strength of student workbook responses	87	68	67	66
17. Percentage of prompted teacher presentation	26	23	33	7
18. Percentage of prompted, student workbook	26	48	72	61
19. Days since two examples were presented	158	24	35	47
20. Total number of examples in program	37	139	18	68
21. Number of student examples on same day as teacher material	3	12	0	21
22. Percent of student examples on same day as teacher material	8	9	0	31
23. Total number of lessons	5	35	10	39
24. Percentage of examples for which correction is specified	0	0	0	0

BEST COPY AVAILABLE

Table 3

First Interview on Main Idea

1. How many lessons teach the topic?
2. What kind of teacher directions are there?
3. What kind of student materials are there?
4. What kind of mastery testing is done?
5. Do student materials appear on same day as teacher materials?
6. Describe the topic as it is taught in the program.
7. Are the materials accurate and complete?
8. Is the presentation adequate for teaching students?
9. Does the material introduce or teach the topic?
10. What things do you think should be changed in the materials?
11. How much time do you think is needed to teach the topic?
12. How often should it be reviewed?
13. How closely do you follow the teacher materials?
14. What things do you usually change?
- 15.* How much time do you usually spend teaching the topic?
16. What percent of the students master the topic?
17. What percent remain basically unchanged?
- 18.* For what percent is the teaching a review of what they know?
19. What part of the teacher materials are the most beneficial?
20. What part of the student materials are the most beneficial?
21. What would a simple example of the topic you might use?
- 22.* What is a more difficult example you might use?
23. How would you teach the student the topic is ____?
24. How would you teach the students the topic is not ____?
25. How would you correct this type of mistake?
26. How would you correct this type of mistake?

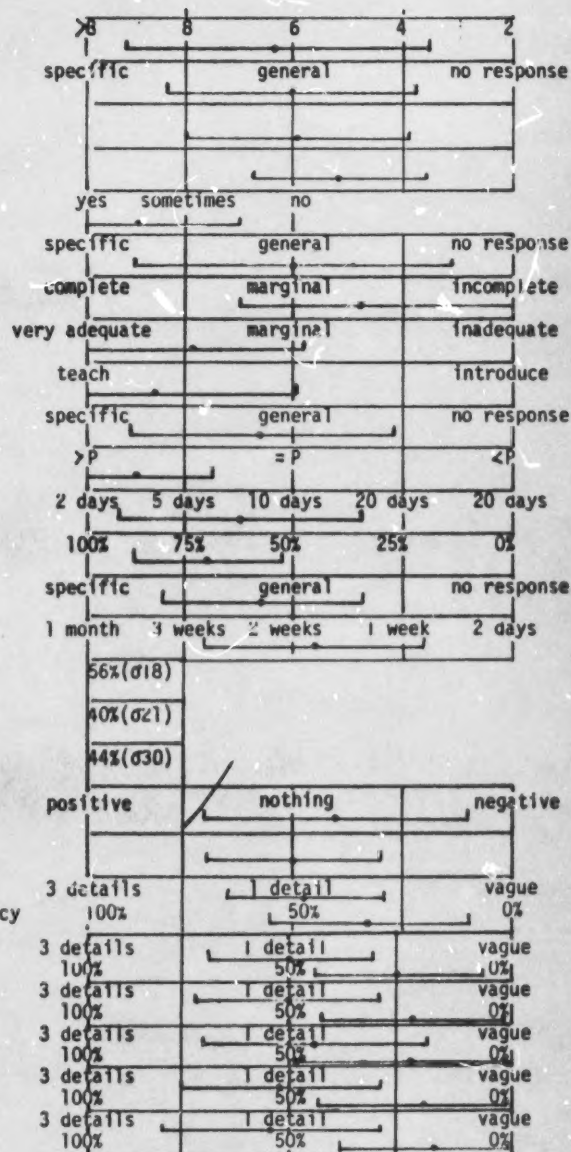
Accuracy

Accuracy

Accuracy

Accuracy

Accuracy

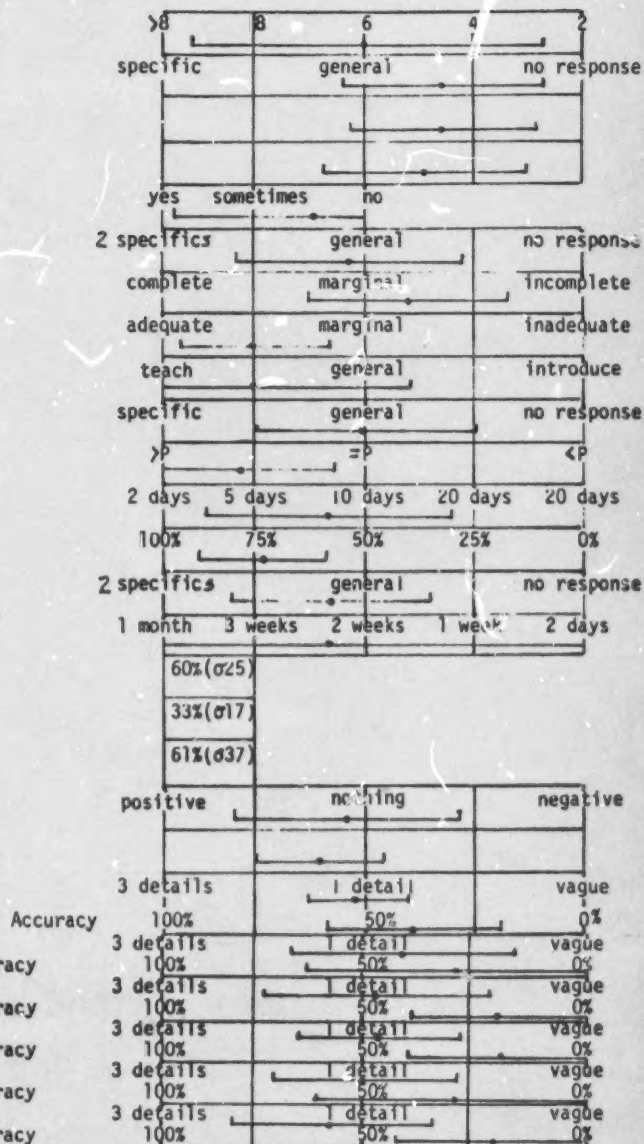


BEST COPY AVAILABLE

Table 4

First Interview on Cause/Effect

1. How many lessons teach the topic?
2. What kind of teacher directions are there?
3. What kind of student materials are there?
4. What kind of mastery testing is done?
- 5.* Do student materials appear on same day as teacher materials?
6. Describe the topic as it is taught in the program.
- 7.* Is the materials accurate and complete?
8. Is the presentation adequate for teaching students?
9. Does the material introduce or teach the topic?
10. What things do you think should be changed in the materials?
11. How much time do you think is needed to teach the topic?
12. How often should it be reviewed?
13. How closely do you follow the teacher materials?
14. What things do you usually change?
- 15.* How much time do you usually spend teaching the topic?
16. What percent of the students master the topic?
- 17.* What percent remain basically unchanged?
- 18.* For what percent is the teaching a review of what they know?
19. What part of the teacher materials are the most beneficial?
20. What part of the student materials are the most beneficial?
21. What would a simple example of the topic you might use?
- 22.* What is a more difficult example you might use?
23. How would you teach the student the topic is this?
24. How would you teach the students the topic is not this?
25. How would you correct this type of mistake?
26. How would you correct this type of mistake?



or the things they do change in the program (item 14). The only average that differs by more than 30% between main idea and cause/effect is item 15, teachers indicate that they spend less time teaching cause/effect.

Questions 16-18 ask teachers about student performance. For both main idea and cause/effect, teachers estimate that slightly over 50% of the students master the topic (56% and 60% respectively). There are differences in the estimation of students that remain unchanged (40% for main idea and 33% for cause/effect) and the percentage of the student who already knew the topic (44% and 61%).

Questions 19 and 20 ask the teachers to name specific parts of the teacher materials and student materials that are particularly beneficial. For main idea, their averages were neutral to mildly negative; in other words, they did not suggest anything that was particularly beneficial. For cause/effect, the averages were slightly positive.

Questions 21-26 ask teachers to describe how they would teach certain items or correct certain errors. For all items on both the topics of main idea and cause/effect, the performance is uniform; teachers average only one detail or specific bit of information per item.

Items 21-26 were rated on accuracy as well as specificity. To receive a score of perfect accuracy on any of these items, a teacher simply had to provide a few of many possible, accurate statements. The teacher's accuracy ratings were very low (below 25%) when asked to identify a simple example of the topic, identify a more difficult example, indicate how to teach positive examples of the topic, indicate how to teach a negative example, and specify how to correct two types of mistakes. The accuracy score was very consistent from item to item, and stable across main idea and cause/effect (varying less than 20% on any item).

Teachers were not able to give many specifics about why they changed program-specified procedures (item 14), nor were they able to be specific or accurate on suggesting ways for teaching a negative example, i.e., "The

main idea is not always the first sentence," (item 24). They suggested that greater concentration of lessons was needed but did not agree about whether the skills meshed with other related skills presented in the program.

Cluster Analysis

A cluster analysis was performed to determine possible intercorrelation between various first-interview responses. All clusters are significant at the .05 level. Four clusters suggest the general relationship that the lower the teacher's opinion of the program, the greater the teacher's verbal competency:

Cluster 1

more accuracy on giving example of topic

fewer numbers of lessons teach main idea

more apt to label the teaching is "introduction"

Cluster 2

more accurate they are on giving examples of topic

more specifics they name as changes in the program

more accurate they are on their given correction procedures for an example

Cluster 3

more accurate they are in explaining how to teach the positive example

more time they spend on topic beyond what program specifies

more accurate they are on example of topic

more specific they are in correcting example error

Cluster 4

more accurate they are in giving example of topic

more specific they are in describing teacher materials

more specific they are in describing how the program teaches main idea

The final cluster suggests that the more competent teachers report some significantly different classroom behaviors:

The higher the teacher's accuracy on teaching positive example

The more specifics the teacher gives on correcting example error

The more time the teacher reports spending teaching main idea

The lower the teacher reported percent of students unchanged

Although these correlations are interesting, they should not be taken as extremely strong predictors. In all, they tend to show that the teacher who is generally knowledgeable about instruction has attended to details that the more naive teacher ignores. This tendency does not predict classroom performance of the teacher or student outcomes, merely verbal behavior.

Second Teacher Interview Results

The second teacher interview covered specific questions about the second video-taped lesson. These questions (Table 5) related to student performance, teaching behaviors, and changes in the specified programs.

Teacher estimates of the percentage of the students who know the skills before the lesson (item 1: 36%), the percentage that mastered the skills (item 2: 72%) and percentage that needed more practice (item 3: 58%) were higher than the corresponding questions on the first teacher interview. The indicated amount of additional practice that 58% of the students needed was fairly low--a little more than one extra week of practice. Teachers estimated the percentage of students who would be able to master any skill at 86% (item 4).

Questions 8, 28 and 29 test responses to general teaching techniques. Teachers were able to give less than one specific statement about how they chose students to answer questions (item 8). Their specificity on when to use prompts was greater (28). They estimated they reinforced an exceptionally high 78% of the student responses (item 29).

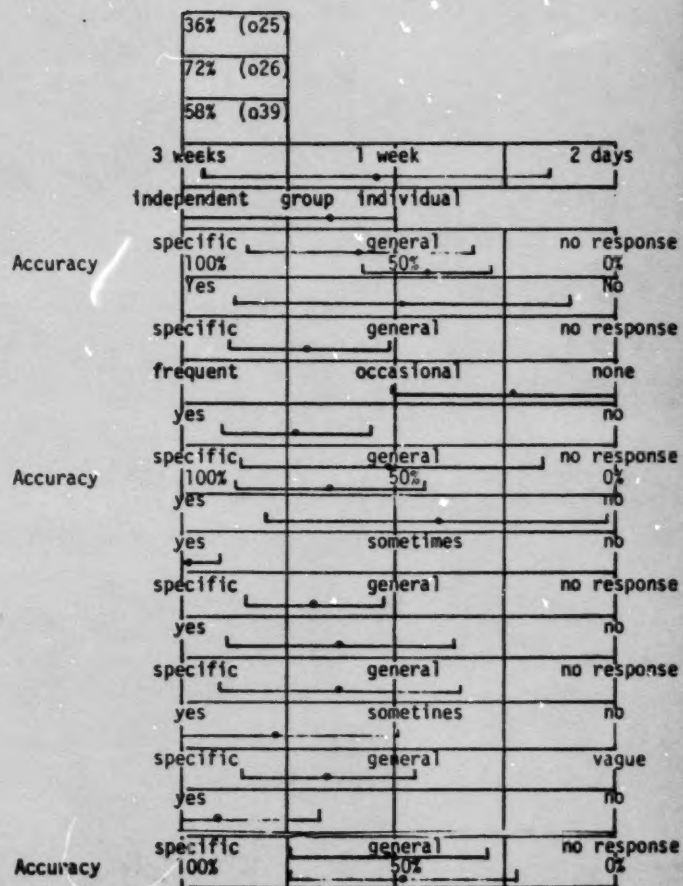
Questions 10-13 relate to interactions between teachers and students.

Table 5

Second Interview on Lesson Taped

1. What percent could do the skills before the lesson?
2. What percent could do the skills after the lesson?
3. What percent need more practice?
4. How much more practice?
5. What would be the most effective type of practice?
6. How would this practice be done?
7. Can students not called on usually answer same question?
8. How do you decide who to call on?
9. What is the frequency of behavior problems in that group?
10. Your rate was _____. Are more responses needed?
11. How could that be accomplished?
12. Is the amount of feedback you receive adequate?
13. Would you follow procedures that give more feedback?
14. You made these changes in the lesson. Why?
15. ____ Percent of the responses were wrong. Is that too high?
16. Why were there so many errors?
17. Would you follow correction procedures?
18. Why is this skill important in the program?
19. Is it a transition step to harder skills?
- 20.* What is that more appropriate task?

* = N<11



BEST COPY AVAILABLE

the same way.

The students in grades 5 and 6 make a larger percentage of errors, which seem to result in a less "positive attitude" on the part of the teacher, who tend to praise a lower percentage of correct responses. This group of correlations suggests that the teaching is less effective in the higher grades (even though what is taught is basically the same as that presented in grade 4).

Student performance. Below is a list of significant and non-significant differences on student performance.

No Significant Difference

1. Percent of students at 50% criterion level

Significant Difference

1. Percent of students at 90% criterion (smaller percentages for 5th and 6th grade students)
2. Percent of students at 75% criterion level (smaller percentages for 5th and 6th grade students)

Although the percentage of students performing at 50% criterion is the same across three grades, a smaller percentage of 5th and 6th grade students perform at the 75% criterion or at the 90% criterion. The trend shows that fewer students in the higher grades have an adequate grasp of the concepts presented by the teachers.

Teacher reports. This list below shows 13 no-difference patterns and 12 significantly different patterns of verbalization.

No Significant Difference

1. Adequacy of presentation
2. Adequacy of feedback from the students.
3. Agreement that more responses are needed
4. Percent of lessons that have both student and teacher materials on same day
5. Agreement that there were more appropriate tasks than the one video-taped

6. Agreement that the topic of main idea was only introduced, not taught
7. Percent of students who needed more practice on video-taped lessons
8. The amount of extra practice these students needed
9. The percent of students who should master any topic in reading
10. Degree to which they follow the program-specified procedures
11. Agreement that they would follow more specific feedback procedures
12. Specificity of how to lower error rate
13. Accuracy of correction procedures

Significant Differences

1. Accuracy and completeness of program (5th and 6th grade teachers rate programs lower)
2. Amount of review the program provides (5th and 6th grade teachers rate program lower)
3. Percent of students who master main idea (5th and 6th grade teachers gave lower estimate)
4. Percent of students who remain unchanged by program (5th and 6th grade teachers gave higher estimate)
5. Percent of students who mastered the video-taped lesson (5th and 6th grade teachers gave lower estimate)
6. Percent of errors was too high (5th and 6th grade teachers showed more agreement)
7. Agreements that they would follow more specific correction procedures (5th and 6th grade teachers showed more agreement)
8. Specific details on how they changed the program (5th and 6th grade teachers gave fewer details)
9. Specific details on how to correct errors (5th and 6th grade teachers gave fewer details)
10. Specific details on when to use prompts (5th and 6th grade teachers gave fewer details)

11. Accuracy of suggestions on when to use prompts (5th and 6th grade teachers were less accurate)
12. Percent of responses they said they reinforce (5th and 6th grade teachers gave lower estimates)

The 5th-6th grade teachers are as vague as the teachers in 4th grade on facts about their program. They agree that main idea is not taught, and they specify roughly the same percentage of students who need additional practice. However, they have a lower opinion of the program they use. They rate the program lower with respect to completeness, review and capacity to teach. 5th-6th grade teachers are more amenable to correction procedures. Finally, the 5th-6th grade teachers seem to be less aware of what they do or why they do it. They are less specific about what they change, when they prompt, and how they correct. In all, their responses convey a greater sense of despair than those of 4th grade teachers.

Comparison of Teacher Reports and Program Analysis Data

Ten items which appear on Table 6 from the teacher interviews have some form of correspondence to program-analysis data. Those items relate to teacher knowledge of the program or teacher judgments about the program. For some judgment items, the teachers' responses suggested knowledge of the program. For the remaining judgment items and the three knowledge items, the teachers responded with little understanding of the program.

There was 73% agreement that more student feedback was needed (item 6) and teachers correspondingly asked $2\frac{1}{2}$ times more questions than were specified in the program. There was a 60% agreement that more appropriate tasks could have been designed (item 10); the programs average: 27% probability of correct interpretation for ten video-taped lessons. There was a 76% agreement that the skills for a topic could be better taught if they were more concentrated (item 7); the programs average 62 days between presentation of two or more examples on main idea. Teachers appear to want more feedback proce-

Table 6
Teacher Reports and Program Data

1. Do student workbook activities for main idea appear on the same day as teacher presentation? Percent of examples specified for same day as teacher presentation.	68% of the time 12%
2. Is what is presented for teaching main idea accurate and complete? Percent of program questions that are ambiguous or not taught. Percent of program questions that are misleading or wrong.	35% 88% 12%
3. How adequate is the presentation for teaching main idea? Probability of correct student interpretation of main idea.	73% 27%
4. Would you say the program teaches or introduced main idea? Percent of examples not taught.	83% introduce 86%
5. How often is the topic of main idea reviewed? Average rate of review.	1 per 8 days 1 per 62 days
6. To what extent should the program have been designed so the students respond more. Percent increase of teacher questions. Average number of questions specified per example of main idea.	73% 251% 2
7. Would the skill be better taught if it was more concentrated? Average days between examples of main idea topic.	76% agreement 62
8. Would you follow procedures to increase the amount of feedback you get from the students? Average number of questions specified per example of main idea. Percent increase of teacher questions.	97% agreement 2 2-251%
9. Would you follow more specified procedures for correcting student errors? Percentage of examples that had 1 or more specified correction procedures.	78% agreement 0
10. Is there a more appropriate task?	60% agreement

Table 8
TEACHER BEHAVIOR DATA BY EXERCISE, TOPIC, AND PROGRAM

	G r a d e	Means by type of exercise			Means by topic taught					Means by program			
		1	2	3	1	3	6	7	11	1	2	3	4
Total percent questions with errors.	27	34	26	19	29	26	28	25	7	18	24	32	25
Total percent questions that were group tests.	16												
Total percent questions that were individual.	84	94	84	84	82	46	94	43	98	97	86	72	60
Total percent errors that were corrected.	37	50	33	18	43	15	47	17	25	20	24	47	46
Total percent errors that were corrected & retested.	10	5	11	8	17	8	7	8	0	1	6	10	30
Total percent of tasks that were models.	20												
Total percent of tasks that were leads.	14												
Total percent of tasks that were models or leads.	34	18	26	7	33	12	26	4	8	11	26	19	30
Total percent of responses that were praised.	46	27	45	62	44	35	27	22	79	72	52	20	41
Total percent of responses that were given general praise.	44												
Total percent of responses that were given specific praise.	2												
Total percent of responses that were given negative feedback.	1												
Rate	4.4	5.4	4.2	4.9	3.3	5.3	5.1	6.2	5.8	4.6	4.0	5.1	3.4
N	29	3	16	5	9	2	2	3	2	4	14	6	3

Type of exercise code 1=mostly reading w/25% or less questioning
2=50/50 reading and questioning
3=most questioning

Topic Code 1=main idea
3=context clues
6=story comprehension
7=vocabulary
11=parts of a book

Program Code 1=Houghton-Mifflin
2=Ginn
3=Scott Foresman
4=Holt

BEST COPY AVAILABLE

3.4 and 6.2 responses per minute). Praising behavior varies widely across topics.

Scott Foresman teachers achieve the highest response rate (5.1), the best overall percentage for praise (20%), and the highest rate of correcting mistakes (47%). To counter these positive features, Scott Foresman achieved the highest error rate (32%). Holt teachers achieve the best percentage of individual tasks (60%), the highest percentage of corrections that involve a retest (30%) but the slowest presentation rate (3.4 tasks per minute). The lowest error rate and the lowest rate of model-led tasks (11%) were achieved by Houghton Mifflin; however, Houghton Mifflin also achieved the highest rate of individual turns (97%) and the lowest rate of specific corrections (1%). Ginn excelled in nothing.

The pattern is somewhat reminiscent of that seen in the program-analysis section. Individual programs vary on different criteria; however, the overall lack of control seems to neutralize the programs and reduce them to the same level.

Table 9 shows the performance of the teachers on two different tapings. Each cell is divided. The top number indicates performance on the first tape, the bottom number indicates performance on the second.

The first five teachers presented the same type of exercises on both the first and second taping. The remaining teachers presented different types on the two tapings.

Two behaviors remain quite stable across all teachers; the rate of task presentation and the percentage of individual children called on. The rate of the teachers does not usually vary more than one-and-a-half tasks per minute, whether the teacher presents the same topic or a different one on both tapings.

The pattern of presenting individual tasks varies more radically; however, if we take into account the variety of activities that were taped, the pattern

Table 9

Teacher Behavior Data on First and Second Video Tapings

Teachers	Rate	% Errors	% Errors Corrected	% Errors Corrected and retested	% Individual responses	% Model and Lead	% Praise
1	6.2 3.8	28 17	53 25	28 0	100 100	3 0	30 75
2	2.5 2.2	0 40	0 83	0 17	93 93	50 7	93 40
3	6.3 5.0	30 50	7 36	2 0	60 100	0 28	0 36
4	6.8 6.9	25 30	46 5	23 0	32 19	5 11	11 24
5	5.3 6.0	6 8	0 50	0 0	97 100	13 2	91 67
6	2.3 3.0	15 13	25 8	0 0	92 100	31 65	54 79
7	5.5 3.3	19 16	0 4	0 0	33 82	8 58	55 58
8	5.2 3.7	9 22	19 25	0 16	84 72	6 13	67 45
9	2.5 2.8	28 49	25 30	16 0	89 94	34 42	56 33
10	3.1 5.2	45 46	60 55	12 0	89 95	71 4	11 27
11	5.5 4.5	23 24	58 57	0 25	100 100	2 23	5 40
12	3.1 2.8	24 23	10 58	10 33	85 43	13 61	58 54
13	4.2 4.2	33 39	36 66	14 25	88 100	49 12	49 29

(For three teachers, one of their two video-tapes was not usable.)

of individual responses remains remarkably stable, varying less than 10% for most teachers.

The other categories deal with responses to students. The pattern for these categories is highly variable from first to second taping. The percentage of errors that are corrected varies more than 25%. The percentage that is retested fluctuates for some teachers. Four teachers, however, consistently do not retest. The model-lead behavior of the teachers also shows a remarkable variation from taping to taping. Apparently, this behavior is in response to content variation, and varies more than 25%. Also, the percentage of responses praised is highly variable, typically varying more than 25%.

The pattern of instability is quite constant across all teachers. Although the first five teachers present the same type of activity on both tapings, their performance is as variable on the response-to-student categories as that of the other teachers. This fact suggests that the variability is not so much a function of the nature of the activity presented as it is a function of perhaps the "mood" or unmeasured aspects of the context.

There are some relationships between the categories. If a teacher corrects more, the teacher praises less. Also, if errors are higher, teachers tend to praise less. (Note that the praise is based on a percentage of correct responses and therefore is not necessarily affected by the number of errors.)

There is also a weak tendency of teachers to praise more frequently when they go faster. The rate of presentation also related slightly to the number of errors and the percentage this is corrected.

The lack of stability of the teachers' performance may be interpreted several ways. Possibly the teachers are trying to show off, or look good; however, since they are not used to correcting and praising, they forget about what they are supposed to do in response to students. While their responses to students vary, the non-response part of their presentation--rate and pattern of calling on students--remains stable.

Table 10

Average Teacher Behavior and Average Program Specifications

1. Total number of questions asked in the program lesson	8
Total number of questions asked by the teacher.	20
Percent more teacher questions over program questions.	251%
2. Percent of program questions that were ambiguous or misleading.	42%
Percent of teacher questions that were ambiguous or misleading.	48%
3. Percent of program questions that were relevant to the topic.	69%
Percent of teacher questions that were relevant to the topic.	24%
4. Student correct interpretation probability from the program.	22%
Student correct interpretation probability from the teacher.	27%
5. Strength of student responses from program questions.	78%
Strength of student responses from teacher questions.	77%
6. Percent of prompted responses in the program.	12%
Percent of prompted responses from the teacher.	22%

Another interpretation is that the teachers use a variety of responses to the students, which are determined by "mood" or other subtle aspects of the situation. The responses, therefore, vary from situation to situation, while the non-response aspects of the presentation remain stable.

Regardless of the interpretation we select, the response behavior of the teachers is apparently not determined by technical, rational considerations but more by caprice or chance. The teachers, in other words, are very poor at responding to students in a consistent manner.

Comparison of Teacher Behavior and Program Analysis

Table 10 compares the teachers' presentations with the test presentations for corresponding lessons. The group of lessons include main idea and context clues. Details of program analysis for these lessons, such as percentage of tasks prompted, do not correspond to those presented in the program-analysis section because the earlier means are based on the entire fourth to sixth grade, rather than just the classes that were video-taped.

As Table 11 indicates, the teachers deviate substantially from the test specifications (69% deviation). The changes they imposed on the text specification consist primarily of additional questions, additional prompts, and changes in the fidelity of the questions.

There is almost no difference between the program and the teacher presentations on the percent of questions that are ambiguous or misleading (item 2), percent probability of correct student interpretation (item 4), or strength of student response (item 5).

The teachers ask $2\frac{1}{2}$ times (251%) the number of questions specified in the program (item 1); however, their rate of irrelevant questions is approximately $2\frac{1}{2}$ times greater than the program average. Their additional questions, in other words, are overwhelmingly irrelevant. Teachers also give about two times more prompts than the programs specify.

The change in fidelity of the teachers' questions is almost completely

Table 11
Lesson Analysis Results

1a. Total number questions, text
1b. Total number questions, teacher
2a. Combined % Amb, Msl, text
2b. Combined % Amb, Msl, teacher
3a. % relevant to concept, text
3b. % relevant to concept, teacher
4a. % probability of correct interpretation, text
4b. % probability of correct interpretation, teacher
5a. % strength of response, text
5b. % strength of response, teacher
6a. % prompted responses, text
6b. % prompted responses, teacher
7a. % variation in number of questions
8a. % variation in content of questions
9a. % deviation from specified procedures
10a. Average percent variation from text

By program				All
Rough- ton- Hiff- lin 2	Ginn	Scott Fors- man	Holt	Mean
15	6		3	8
39	15		7	20
50	47		50	50
45	47		56	43
35	71		100	69
19	21		39	24
21	22		25	22
16	29		33	27
72	90		67	78
90	30		57	77
10	25		0	12
17	21		19	22
167	119		134	131
20	26		32	26
20	48	30	20	28
64	68		76	69

BEST COPY AVAILABLE

accounted for by the fact that teachers add irrelevant questions.

A summary point of great importance. On critical technical details, the teachers do not perform any better than their instructional programs. Their programs, in other words, apparently function as limiters or models on these critical features. The instructional program exhibits no apparent concern for constructing sets of examples that rule out possible fallacious interpretations of the concepts. The teachers' presentation shows this same cavalier attitude about the examples. The instructional program provides spurious prompts, presents great variation in task wording, does not teach minimally-different discriminations, and does not specify corrections. The teachers perform in a parallel manner. They prompt, present a great deal of ambiguous and irrelevant material, fail to teach minimum differences and generally do not correct.

The extent to which the instructional material causes or promotes such teacher behavior cannot be answered by the present study; however, from a purely rational standpoint, it seems reasonable that teachers shouldn't be expected to correct appropriately, present examples that are consistent with a single interpretation, and control other program-related details, if the instructional program they use shows no concern for these details.

Comparison of Teacher Behavior and Teacher Reports

Table 12 presents eight items that compare the teachers' teaching behavior with their verbal reports of behavior. The overall summary of the difference is that teachers do not seem to be aware of precisely what they do when they teach (items 1 and 2). They recognize a need for more student responses, better feedback and lower error rate (items 3, 4 and 5). However, they cannot describe what they actually do when they correct, lead or model, reinforce, or change the procedures of the program (items 6, 7 and 8).

Great discrepancies are apparent between actual behavior and verbal reports about how closely teachers follow the program and what percentage of students

Table 12

Teacher Behavior and Teacher Reports

1. How closely do you follow the program specified procedures for teaching main idea? Average total percent of following program procedures.	70% 31%
2. What percent of the students responses do you reinforce? Percent of responses teacher reinforced.	78% 47%
3. Do you think that the program should have been designed so that students answered more often in the lesson we taped? Percent increase over number of questions asked in program.	73% agreement 251%
4. Do you receive adequate feedback from the students? Average rate of student responses. Average percent of incorrect responses.	20% less than adequate 4.2 response per/min 27%
5. Was the rate of student errors too high in the lesson we taped? Average percent of student errors. Percentage of errors corrected.	62% 27% 37% corrected
6. How and when do you correct student errors? (100% is equivalent to naming 2 specifics) (100% is equivalent to naming 2 accurate points) Percent of errors corrected Percent of errors corrected and then retested.	66% specific 37% accurate 37% corrected 10% corrected and retested
7. When is it appropriate to help or lead the students in a task? specificity (100% is equivalent to naming 2 specifics) accuracy (100% is equivalent to naming 2 accurate points) Percent of tasks modeled or lead	75% specific 54% accurate 31% modeled or lead
8. What things do you change in the way you teach main idea? (100% is equivalent to naming 2 or more specifics) Percent deviation in number of questions asked, percent deviation in fidelity of questions asked. Percent deviation in program specified procedures.	58% specific 151% 26% 28%

they reinforce (items 1 and 2). The teachers' estimates miss by 39% for following the program and 31% for reinforcing.

The teachers tend to agree with the statement that the program should have been designed so that students answered more often (item 3). The teachers showed a possible answer to this shortcoming--an increase in the number of questions asked. However, the additional questions (as noted earlier) are overwhelmingly irrelevant to the concept being taught.

The teachers generally agreed that more feedback was needed from students (item 4); however, their teaching discloses no apparent response to this problem. The teachers produced student responses at a very low rate and presented 85% of the tasks to individuals.

The teachers generally agreed that the rate of student errors was too high; however, the most universal response to errors was not to correct them but to ignore them. (Only 37% were corrected.)

Items 3, 4 and 5 suggest that the teachers lack technical skill. They recognized the problems of student response, feedback and corrections; however, their attempts to solve these problems were abortive and sometimes ritualistic.

The message conveyed by items 6, 7 and 8 is that the teachers apparently do not attend to the details of what they do when they teach. To score 100% specific for any of these items, the teachers were required to name two specific details associated with correcting, leading, or changing the program. Similarly, to score 100% accurate, teachers were required only to make two observations that are true about correcting, leading, or changing the program. The teacher could have said four falsehoods so long as she mentioned two accurate details--any two from a set of possibly hundreds. The teachers performed poorly on these verbal tasks. Their teaching behavior suggests that they may not generally operate from a plan to respond to the student or to the program. Instead, their performance seems more consistent with the idea that they ad lib. If a student catches them off guard by making a

mistake, they go to another student. If the student seems to have trouble producing a response, they give "clues" or prompts. If the content of the material presented by the text prompts an association in the teacher's mind, it may appear as an additional question to those specified by the program-- a question that is probably irrelevant to the concept and one that serves to pull the students farther from an articulate understanding of what the test is trying to convey.

STUDENT PERFORMANCE RESULTS

After the second taping, students were presented with test worksheets covering the material presented during the taped lesson. These worksheets provided a fair test of the concepts that were supposed to be taught.

Table 13 presents a summary of student performance on the eight topics that were video-taped. The columns show the percentage of students who perform at 90% or above, 75% or above, and 50% or above.

The trends are fairly stable across different topics. Except for relevant details, and map skills, the percentage of students who achieve 90% or higher on the test sheets ranges between 0% and 15%. The mean for 90% criterion is 12% of all students.

The trend is even more stable for the 75% criterion. An average of 30% of all students score 75% correct or better on the test items. The extremes are context clues, for which 0% scored 75% or better and relevant details with 82% of the students achieving 75% correct.

An average of 55% of all students score at least 50% correct, which means that nearly half of the students tested were not able to correctly respond to half the text items. Relevant details and context clues are again the extremes. The remaining topics range between 56% of the students and 70% of the students.

Although more tapes were made of main idea than of any other topic, main idea seems to be a near average topic, deviating from the mean no more than three percentage points in the three criterion columns.

Although there is some variability across topics, conclusions suggested by Table 13 are confirmed here: students are not taught by the presentations, which means that the presentations are not designed to teach. The topic of main idea is taught by all programs; however, only one third of the students achieve a criterion of 75% correct, which is modest. Other topics, which are typically taught in all programs--fact and opinion, context clues,

Table 13
Criterion Percents by Topics

Topic	Mean percent of students at or above level			Programs tested
	90% criterion	75% criterion	50% criterion	
Main idea (8)	10%	33%	58%	HM,G SF,H
Key Words (3)	8%	32%	65%	HM,G
Map Skills (3)	30%	33%	56%	HM,G
Inferences (2)	15%	30%	62%	G
Context Clues (6)	0%	0%	15%	G,SF H
Relevant details (2)	24%	82%	99%	HM,G
Cause effect (1)	10%	30%	60%	G
Fact/opinion (1)	0%	25%	70%	SF
Means across all topics	12%	30%	55%	

cause/effect--perform no better.

Table 14 shows the criterion levels achieved by different instructional series. Houghton Mifflin and Holt achieve the lowest percentage of students who perform at the level of 50% correct (48% and 44%). Scott Foresman achieves the lowest percentage for the 90% criterion and the 75% criterion.

Although there is variability across programs, the percentage of students who achieve either 90% correct or 75% correct is very low, with no program exhibiting even minimally acceptable performance.

Table 14
Criter Percents by Program

Mean percent of students at level			# of tests in program
90%	75%	50%	
6%	30%	48%	6
14%	36%	61%	12
5%	23%	58%	4
12%	28%	44%	4
12%	30%	55%	

Comparison of Student Performance and Program Analysis

The problem with trying to compare student performance with the program is that the program does not interact directly with the students. The program is mediated by the teacher's presentation which means that the students receive the teacher's interpretation of the program specifications. Therefore, the comparison of program and student performance must be prefaced by the assumption that certain outcomes are predicted if the teacher actually follows the program or retains those aspects of the program on which the outcome is based.

We can assume that the program will determine at least part of the communication between teacher and student. The workbook material will probably be the only form of independent material the students receive. Therefore, the weakness of the workbook should be at least partially responsible for student performance. The workbooks spuriously prompted 49% of the examples in the taped lessons. They presented primarily multiple-choice items. Twenty-five percent of the presentation consisted of visual distraction.

These characteristics would predict poor student performance unless they were preceded by very strong teacher presentations.

The overall characteristic of the program that might have the greatest predictive value is the probability of correct interpretation. If the teachers communicated in a way that led to four or more possible interpretations and if the students did not know the concept, it is highly improbable that more than one-fourth of them would learn the concept from the presentation. This assertion, however, is an oversimplification. If the amount of practice provided on the topics were multiplied by a large number, students would, through practice and feedback on performance, learn the concepts. Unfortunately, the amount of practice provided by the programs is probably not great enough to assure that one of four students will master the typical topic.

Comparison of Student Performance and Teacher Reports

The study provides two different samples of student behavior; performance during the lesson and performance on the test sheets.

Performance on the lessons discloses that:

Students make mistakes at a rate of 27%

Even modest corrections are provided on only 37% of the mistakes

Articulate corrections that involve a retest are provided on only 10% of the errors

These facts correspond to items in the teachers' report. Teachers indicated that they generally thought the rate of errors was too high (65% agreement with the idea that errors were too high). They rated the feedback they received from students as slightly less than adequate (9% less than adequate); and they endorsed the idea that the instructional program should be designed to provide more feedback and more specific correction procedures. These responses suggest that the teachers recognize the problems of student mastery and some of the more obvious problems associated with the presentation.

Table 15 provides additional comparisons. The first four items are based on the second interview and cover a variety of topics. Items 5-7 are based on the first teacher interview and deal only with main idea.

According to item one, teachers set the ideal level of mastery quite high, contending that 86% of the students should master any reading skill. Whether we use 90% correct or 75% correct as the criterion for "mastery", the actual performance of the students is at least 50% from the teacher's ideal.

The discrepancy of item one is paralleled by item 2. Teachers judged that 72% of the students could do the workbook exercises after the lesson; however, only 7% achieved the 90% criterion on the worksheets that tested the lesson, and only 30% achieved the 75% criterion.

The teachers judge that 58% of the students need more practice on the topic. Seventy percent performed below the 75% criterion, however. Teachers

Table 15

Student Performance and Teacher Reports

Main Idea

1. What percent of the students should master any reading skill?	86%
Percent of students at 90% criterion on all topics.	12%
Percent of students at 75% criterion on all topics.	30%
2. What percent of the students could do the workbook exercises after the lesson was taped?	72%
Percent of students at 90% criterion level on all topics.	12%
3. What percent of the students need more practice on the topic taught?	58%
Percent of students below 75% criterion level on all topics.	70%
4. How much practice do they need?	1 week
Percent of students below 50% criterion level on all topics.	55%
5. What percent of the students master main idea?	56%
Percent of students at 90% criterion on main idea	10%
Percent of students at 75% criterion on main idea.	33%
6. What percent of the students remain unchanged?	40%
Percent of students below 75% criterion on main idea.	67%
7. How deficient is the program for teaching students main idea?	16%
Percent of students below 75% criterion on main idea.	67%

Table 16
Student Performance and Teacher Behavior

	Ideal Percent	Observed Teacher Percent	Deviation
1. Percent of questions ambiguous or misleading	0%	48%	48%
2. Percent of questions relevant to topic	100%	24%	76%
3. Percent probability of correct interpretation	100%	27%	73%
4. Percent strength of student response	100%	77%	23%
5. Percent of responses prompted	0%	22%	22%
6. Percent errors corrected	100%	37%	63%
7. Percent of questions with errors	0-12%	27%	
8. Percent of questions with individual responses	40-75%	85%	
9. Percent of tasks with models or leads	0-25%	31%	
Average Deviation			51%
Percent below 50%			42%

relate the poorness of the programs to the poorness of the student performance.

Table 16 once more compares the ideal teacher behavior with the behavior achieved by the teachers. Item 1-6 have the ideal of either 0% or 100%. There is a range of possible variation for the other items.

The teachers' average deviation from the ideal is 51%. Interestingly, this number corresponds closely to the percentage of students who are at the 50% criterion--58%. Other possible relationships may be drawn from Table 16 because so many of the low scores could serve as possible "predictors." Analytically, it seems that the most serious deviations would serve as the best predictors. If the desired interpretation is only about one of four possible interpretations and if only 24% of the questions are relevant to the topic, incredible mislearning is implied by the presentation. Although we cannot put a precise numerical value on it, we would predict that very few students who did not know the concept before the presentation could possibly learn it from the lessons. This prediction is made for two reasons: because the presentations (1) imply that irrelevant facts are somehow relevant, and (2) permit patterns of responses that are correct but that are consistent with interpretations other than the intended ones.

SUMMARY

Four major categories were examined--the instructional programs, the teachers' verbal reports of teaching, actual teacher behavior, and student outcomes. The results are consistent.

1. The programs are not designed with apparent regard to the technical details of teaching.
2. Although teachers can generally identify serious teaching problems, the teachers' verbal reports are unreliable; void of detail and generally low in accuracy.
3. The teaching performance of teachers varies in some detail, but consistently fails to communicate information about the concept, secure adequate

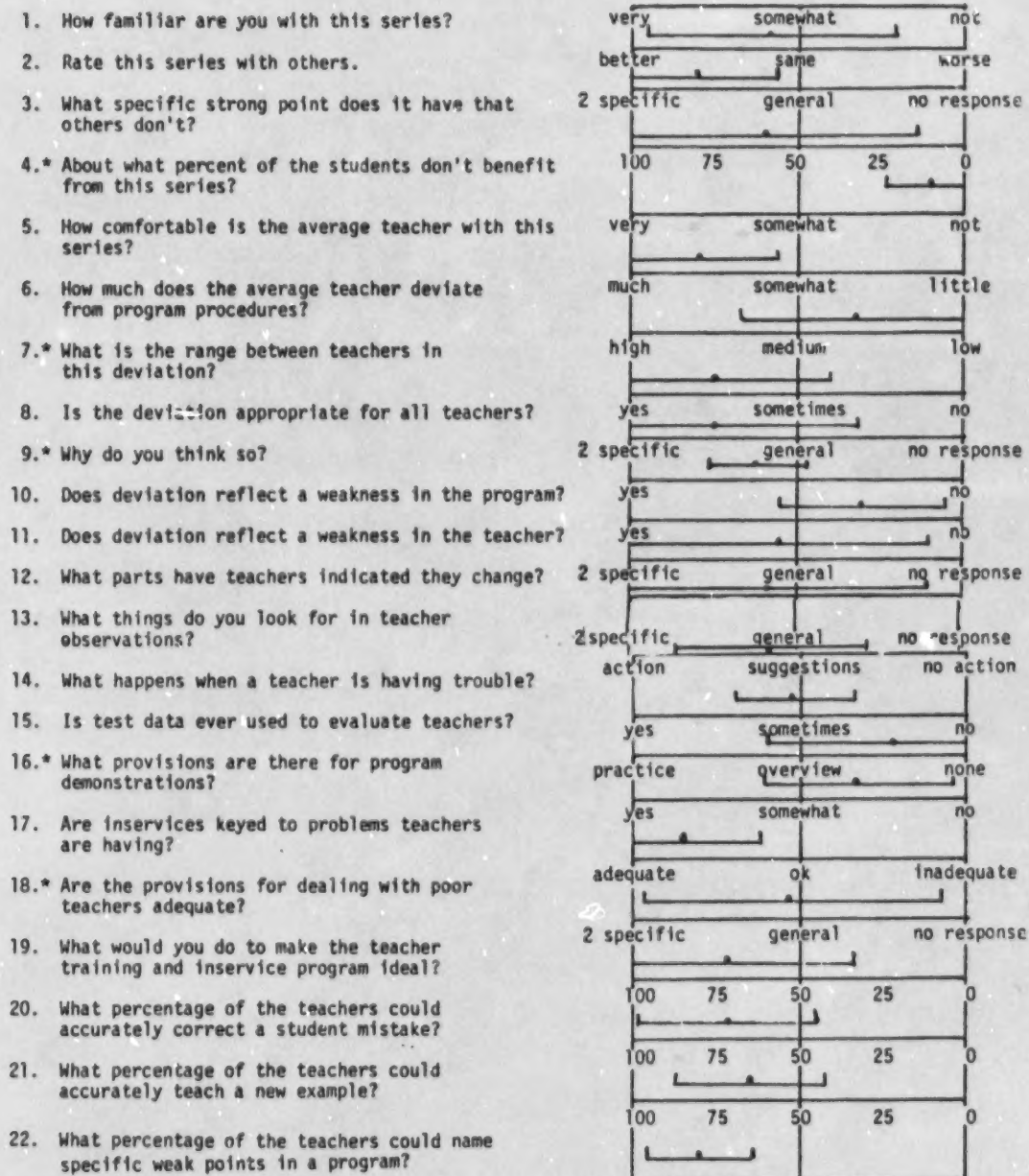
feedback, or firm students.

4. The student outcomes are consistent with the weak programs and weak teaching--with nearly half the students failing to perform at the 50% criterion level of performance and only one-third of the students achieving the 75% criterion.

The results are stable across teachers, across instructional programs, across grade levels, and across topics.

The teachers in grades 4, 5 and 6 seem to be caught in a trap. The programs are not capable of teaching nor of training teachers in what should be done. Tasks are not reviewed on a reasonable schedule, and what should be "presented" (according to program specifications) is not taught in the allotted time. The teachers are therefore reinforced for not teaching and for dealing in irrelevancies. The teachers remain only vaguely aware of what they do when they teach because they do not receive feedback about the importance of attending to particular details. As a result, the teachers' verbal reports are not reliable, either about the program or about student performance. The students, meanwhile, fail to learn the skills that are presented, and their level of failure is both high and consistent.

Table 18
Administrator Responses



* = N < 5

ADMINISTRATOR INTERVIEWS

During the period of April 12 to May 17, 1979, administrators in three sites were interviewed. Table 17 lists the administrators, their position and community.

Table 17

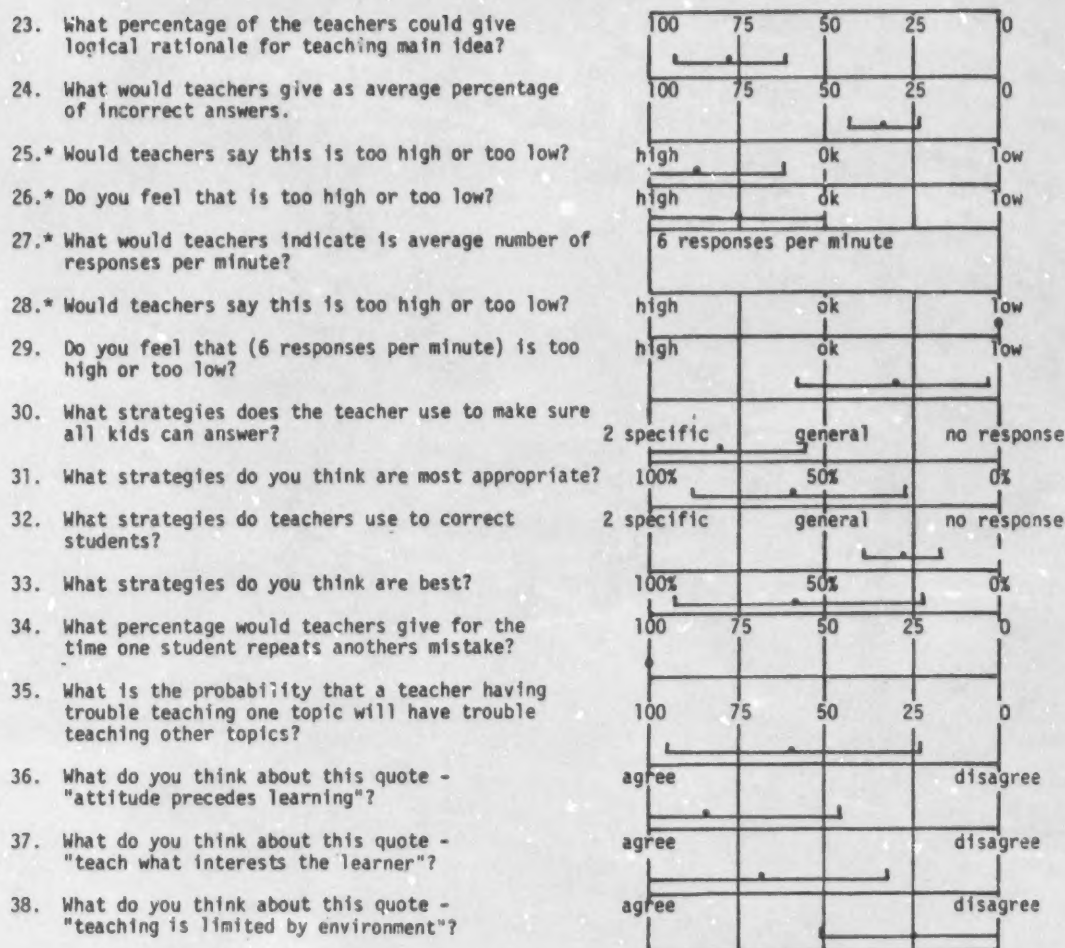
<u>Administrator</u>	<u>Position</u>	<u>Community</u>
1	Principal of School	Bridgeport
2	Reading Supervisor for School	Bridgeport
3	Supervisory Elementary Instruction	Bridgeport
4	Principal of School	Dayton
5	District Reading Supervisor	Dayton
6	Director of Elementary Education for District	Dayton
7	Principal of School	Sacramento
8	Reading Supervisor for Three Schools	Sacramento

Each interview lasted approximately 15 minutes and consisted of 47 items. The interviews attempted to determine the administrators' knowledge of a) the instructional program used by the teachers, b) teacher behavior, c) teachers' opinions and judgments about the program and about teaching, and d) appropriate teaching and program strategies.

Results

Questions 1-4 in Table 18 indicate the administrators knowledge of the program. Administrators judge themselves to be reasonably familiar with the instructional program (item 1) but were able to give little more than general statements about the program's specific strong points (item 3). They rate the program they use considerably better than other programs (item 2), but still estimate that an average of 13 percent of the

Table 18 (Continued)



* = N < 5

students do not benefit from the instructional program (item 4).

Eight items ask about teacher behavior (5, 6, 7, 8, 11, 28, 30, 32). Administrators believe the average teacher is quite comfortable with the program (item 5) and does not deviate a great deal from the program-specified procedures (item 6). This deviation is appropriate for most of the teachers (item 8), but generally reflects a weakness of the teacher (item 11) rather than a weakness of the program (item 10). The administrators rate their teachers' competencies at about a C+ average (items 20-24). They estimate 71% of the teachers can give adequate correction procedures; 68% can give accurate procedures for teaching new examples; 82% can name specific weakpoints in the instructional program; 79% can give a logical rationale for teaching main idea.

Questions 24-33 ask the administrator to estimate what the teachers would say about different issues and then to respond with a personal evaluation of each issue. For item 24-26, administrators estimate that teachers would judge 37% of the responses in a lesson to be wrong, that the teachers would say that this rate is too high, and administrators tend also to agree that it would be too high. In response to questions 27-29, administrators predict the teachers would give six responses per minute as the average rate during a lesson, and judge that the teachers would feel more strongly about this rate being too low than they do. In identifying strategies that the teachers use to insure that all students can answer the questions (item 30 and 31), the administrators show fairly good specificity, their accuracy is slightly above 50%. In identifying correction procedure strategies that the teachers use (item 32 and 33), the administrators are vague and slightly more than 50% accurate. In response to items 36-38, administrators exhibit the same general pattern of responses as teachers, generally endorsing the first two quotes and not endorsing the third.

Comparison of Administrators' Judgments with Teachers' Behavior

Four items provide a comparison of administrators' judgment of teacher behavior and the actual behavior. These items appear in Table 19.

As Table 19 shows, the administrators' judgment errs in the direction of attributing more skills to teachers. They misjudge the percentage of students who apparently do not benefit from the program, the percentage of deviation from the program, the percentage of errors, and the rate of teacher presentation. Only on the percentage of incorrect responses, however, do the administrators err in the direction of over-estimating the problem. For all other responses, the judged behavior is not as good as the observed behavior.

Table 19

Comparison of Administrators' Judgment and Observed Behavior

<u>Item</u>	<u>Judged Value</u>	<u>Actual Value</u>
1. Percentage of students who do not benefit from program	13%	55% below 50% correct
2. Percentage of teacher deviation from program specifications	33%	69%
3. Average percentage of incorrect responses	37%	26% errors during Teacher Presentation 50% errors average for worksheets
4. Average student response rate during lessons	6 per minute	4.4 per minute

The administrators seem to be more accurate about the most obvious aspects of teacher-student interaction; the rate of errors and the rate of the teachers' presentation.

Comparison of Administrators' and Teachers' Verbal Reports

Table 20 presents ten items that compare the judgments of the administrators with those of the teachers.

The administrators' judgment of the teachers' ability to explain appropriate procedures for teaching new examples overestimated the teachers' skills. Only 15% of the teachers scored at the 75% accuracy level or above. The average accuracy of the teachers is only 33%. The administrators' judgment of the teachers' abilities to provide a logical rationale for teaching main idea also overestimated the teachers' specificity and accuracy. Only five teachers scored above the 75% specificity in describing the program's teaching of main idea. The administrators were 87% in agreement with the notion that the teachers would judge their error rate too high. The teachers expressed less agreement - 62%. A similar pattern occurs on items 4 and 5. The administrators judge the teachers to be more self-critical than they actually are.

The administrators give better answers than the teachers for items 6 and 9. The administrators provide a more specific account of strategies for assuring that all students answer the teachers questions. Administrators also provide more accurate descriptions of appropriate correction procedures.

Administrators apparently believe that teachers are more conscientious and critical of their teaching than they actually are. Administrators give some explanations of teaching that are either more specific and accurate than those provided by the teachers. Administrators are apparently aware of two aspects of teaching--1) what teachers say, and 2) the more visible aspects of the program and the teacher-student interactions. At the same time, they do not rate the skill level of teachers very high. They think that the chances are pretty good (over 50%) that if the teacher has trouble teaching one topic, the teacher will have trouble with other topics. They express the idea that less than 2/3 of the teachers can give a logical explanation of why main idea is taught or can provide precise teaching. They are able to give some specific details about what would be involved in an effective pre-inservice program.

Table 2C

Comparison of Administrators' and Teachers' Responses

	Administrators' Judgement	Teacher Comparison
1. What percent of the teachers could explain appropriate procedures for teaching new examples?	63%	15% of teachers who were 75% accurate 33% accuracy
2. What percent of the teachers could give a logical rational for teaching main idea?	75%	36% of teachers who were 75% specific 65% specific
3. Would teachers say error rate of students is too high?	37%	26% errors
4. Would teachers say that their presentation response is too low?	87% agreement	62% agreement (25% observed error rate)
5. Do you think that is too low?	70% agreement	
6. What strategies does the teacher use so that all students can answer the questions?	81% specific	65% specific (accuracy)
7. Which strategies do you think are most appropriate?	60% accurate	40% accurate
8. What strategies do the teachers use to correct errors?	30% specific	65% specific (accuracy)
9. What strategies do you think are most appropriate?	50% accurate	41% accurate
10. What percent of the teachers could accurately tell how many lessons in the program taught main idea?	59%	50% above 75% accuracy

Clusters from Administrators' Interviews

A high percentage of the questions presented to the administrators ask about their judgment of what teachers do, information that the administrators may not have. Because of the small number of items for the administrator questionnaire, the procedure used for determining clusters is different from that used for determining teacher correlations. All teacher correlations are intercorrelations. For the administrator clusters, the first mentioned item correlates with each of the other items; however, the various items do not necessarily intercorrelate (item 3 does not correlate with 1, 2, and 5, for example).

Two major clusters were distilled from the administrators' interviews. Cluster 1 suggests that the administrator who is more specific in dealing with management has a lower opinion of teachers' competency, and of provisions for dealing with weak teachers.

Cluster 1

1. The more specifics he can name for ideal management situations
2. The more he believes deviation is the result of weak teaching (.63)
3. The less realistic he rates the provisions for dealing with weak teaching (.69)
4. The lower his estimated percentage of teachers who can adequately correct errors (.62)
5. The lower his estimated percentage of teachers who can adequately teach new example (.64)
6. The higher his estimate of student errors (.71)

Cluster 2 suggests that the administrator who is unable to provide specific details about ideal management has a much higher estimation of teachers. This administrator is able to name fewer specific details of the series used in the schools.

Cluster 2

1. The more the administrator believes teachers can explain rationale for teaching main idea
2. The fewer specifics he can name for idea management situation (.64)
3. The more comfortable he thinks the teachers are (.69)
4. The more he believes deviation from program is appropriate for all (.84)
5. The higher the percent for teachers who can adequately correct errors (.78)
6. The fewer specifics he can name as strong points of the program used (.58)
7. The more they give specifics on program strong points, the higher they rate the series (.89)

In addition to the clusters, a strong correlation (.89) exists between the number of specifics administrators give about the program's strong points, and how highly they rate the program with the higher rating correlated with more specific knowledge.

QUESTIONNAIRE

Procedures

A questionnaire was developed to determine the extent to which a larger sample of teachers responded in a manner similar to that of the teachers in the primary study. Accordingly, this questionnaire contained items about program specification, teaching behavior, and student outcomes. Additional items provided biographical information about the teachers. The questionnaire provided a total of 94 scorable responses.

During the week of April 23, 1979, the questionnaire was mailed to 3085 teachers in grades 4, 5 and 6. The mailing list, prepared by Market Data Retrieval, provided a random sample of teachers balanced for size of school, location of school, and public/parochial types.

Results

From the acceptable returns, received by May 23, 1979, 493 (16%) returns were used in the analysis.

Questionnaires were analyzed to determine means, standard deviations, correlations, cluster identification, grade level trends, and comparison of questionnaire responses with responses of teachers in the study.

Responses to the questionnaire disclose that items are remarkably independent of other items. Only eight correlations above .40 occurred among the more than 1000 item correlations. This vast majority of correlations were virtually zero, indicating that the vast majority of items measure something different from what is measured by any other item. (See appendix VIII.)

Table 21 presents means and standard deviations for responses to all questionnaire items except those involving non-choice responses (such as indicating the program that is used).

According to the questionnaire, the classroom is average and the teacher spends nearly six hours a week on reading instruction (Section 1, 5, and 8).

BEST COPY AVAILABLE

10. How often is the topic of the main idea reviewed? 2 days 5 days 20 days 40 days
11. How often do you think main idea should be reviewed? 2 days 5 days 20 days 40 days

Section IV

1. Please rate the teaching difficulty of the following reading comprehension topics.

main idea

cause/effect

implied details

vocabulary

outlining

reference skills

reading for purpose

following directions

context clues

drawing conclusions

deductions

determining mood

point of view

fact versus opinion

determining plot

analogies

sequence of events

predicting outcomes

easy

difficult

agree

disagree

2. "... the school experience must help [students] develop positive attitudes toward themselves... This in turn provides a base on which the programs' other [academic] projects can be built."
3. "... learning is likely to be more effective if it grows out of what interests the learner..."
4. "... schools bring little influence to bear on a child's achievement that is independent of [his/her] background and social context..."

The teacher generally follows about two-thirds of the program's specification (Section II; 1) but follows 76% of specified directions for main idea (Section III; 7). Teachers judge that 18% of the student responses will be wrong (Section II; 4), that 81% of the students should master any particular topic taught (Section II; 5), and that the program instructions tend to be specific (Section II; 12). The average teacher corrects 68% of the mistakes (Section II; 6), most frequently by calling on another student or by telling the responding student that the answer is wrong and then repeating the questions (Section II; 7). The teachers call on individual students 70% of the time (Section II; 9) and reinforce 72% of the correct responses (Section II; 8).

The teacher judges that on 52% of the lessons, worksheet exercises appear on the same day as teacher-directed activities for a topic (Section II; 10). She judges that 32% of the students have the concept of main idea before it is taught (Section III; 3) and that 60% master main idea (Section III; 4). The students reportedly have a variety of problems with main idea, the least of which is confusing the main idea with the title (Section III; 6). The more serious problems include confusing main idea with the first sentence, supposing that main idea must be a sentence from the passage, being unable to deal with main ideas that do not contain words from the passage (Section III; 6).

When teaching main idea, the teacher reports that she skips 20% of the specified activities, adds 34%, and changes 26% (Section III; 8). She judges that the topic of main idea is reviewed every 11 days but that it should be reviewed every 5 days (Section III; 10 and 11).

In comparing various comprehension topics, the teacher rates vocabulary as the most difficult to teach, followed by fact versus opinion, reading for purpose, reference skills, and context clues (Section IV; 1). Main idea is rated as a relatively easy-to-teach skill (2.2 difficulty compared to 3.3

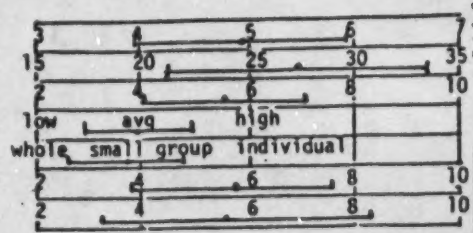
ORIGINAL?

481

81

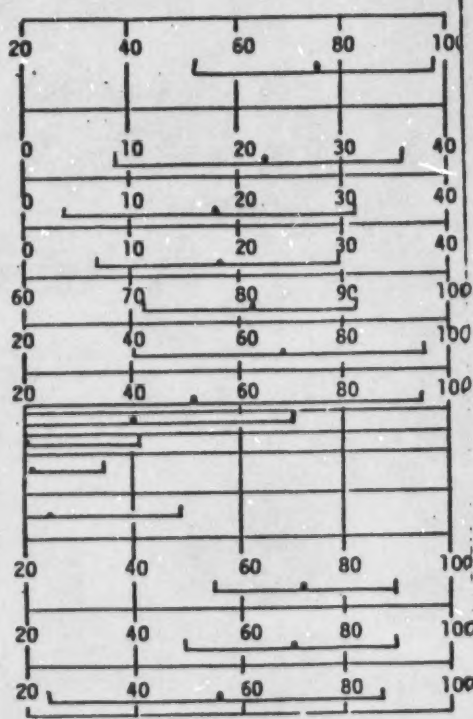
TABLE 21 - QUESTIONNAIRE RESPONSES
SECTION I

1. What reading program do you use? (If you use more than one, list the one you use most and answer the questions based on that program.)
2. What grade do you teach?
3. How many students do you have for reading?
4. How many years teaching experience do you have?
5. How would you describe your reading class?
6. How would you describe the teaching instruction in your class?
7. How many hours per week do the students spend in reading instruction?
8. How many hours per week do you spend actually teaching reading?



Section II

1. In oral teacher directed activities how closely do you follow the procedures that are specified in the teacher's guide?
2. If a good student makes a mistake on an orally-presented reading activity, what percentage of the time will at least one other student repeat the same mistake?
3. If a low student makes a mistake, what percentage of the time will at least one other student repeat the same mistake?
4. During an average lesson, what percentage of the student responses will be incorrect?
5. As a generally acceptable guide, what percentage of the students should be able to master any particular skill in this reading program?
6. What percentage of the mistakes made during oral presentations do you correct?
7. Indicate the percentage of time you use each of these corrections:
 - Call on another student
 - Tell the student the answer is wrong and repeat the question
 - Tell the answer and then repeat the question
 - Tell the answer, repeat the question, then repeat the question to the whole group
 - Ignore the mistake and permit the students to discover a natural consequence of their error
8. Most reading programs make general comments about praising and reinforcing students for correct responses. What percentage of the student responses do you reinforce in some way?
9. What percentage of the time do you call on individual students to answer questions (rather than the whole group or class)?
10. In what percentage of the lessons do worksheet exercises for a topic appear the same day as teacher activities for that topic?
11. There are places in the program where it says to help or guide the students. In what situations is this appropriate? (Check all that apply)



- Students have just learned the skill
- Students rarely use the skill and tend to forget it
- Students need the help or they will never master the skill
- Students have never really been taught the skill
- Students must answer an open-ended question

79%	(SD 40)
55%	(SD 40)
59%	(SD 48)
54%	(SD 39)
35%	(SD 41)

12. How would you describe the teacher directions of your program?
 - General statements of what to do
 - Specific steps you are to follow
 - Specific steps you are to follow and some direction on what to say
 - Detailed scripts of exactly what to say and do

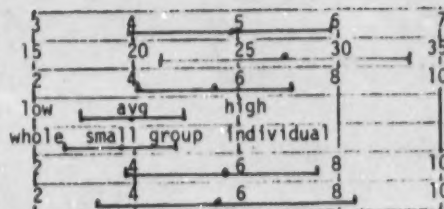
19%	
9%	
47%	
15%	

BEST COPY AVAILABLE

BEST COPY AVAILABLE

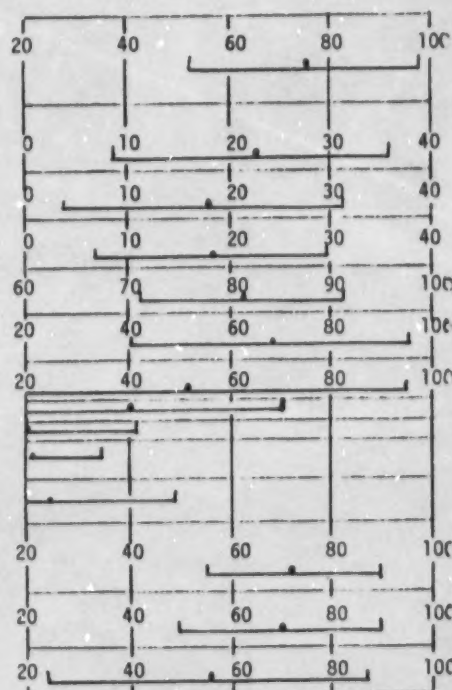
Section I

1. What reading program do you use? (If you use more than one, list the one you use most and answer the questions based on that program.)
2. What grade do you teach?
3. How many students do you have for reading?
4. How many years teaching experience do you have?
5. How would you describe your reading class?
6. How would you describe the teaching instruction in your class?
7. How many hours per week do the students spend in reading instruction?
8. How many hours per week do you spend actually teaching reading?



Section II

1. In oral teacher directed activities how closely do you follow the procedures that are specified in the teacher's guide?
2. If a good student makes a mistake on an orally-presented reading activity, what percentage of the time will at least one other student repeat the same mistake?
3. If a low student makes a mistake, what percentage of the time will at least one other student repeat the same mistake?
4. During an average lesson, what percentage of the student responses will be incorrect?
5. As a generally acceptable guide, what percentage of the students should be able to master any particular skill in this reading program?
6. What percentage of the mistakes made during oral presentations do you correct?
7. Indicate the percentage of time you use each of these corrections:
 - Call on another student
 - Tell the student the answer is wrong and repeat the question
 - Tell the answer and then repeat the question
 - Tell the answer, repeat the question, then repeat the question to the whole group
 - Ignore the mistake and permit the students to discover a natural consequence of their error
8. Most reading programs make general comments about praising and reinforcing students for correct responses. What percentage of the student responses do you reinforce in some way?
9. What percentage of the time do you call on individual students to answer questions (rather than the whole group or class)?
10. In what percentage of the lessons do worksheet exercises for a topic appear the same day as teacher activities for that topic?
11. There are places in the program where it says to help or guide the students. In what situations is this appropriate? (Check all that apply)
12. How would you describe the teacher directions of your program?



79%
55%
59%
54%
35%

19%
19%
47%
15%

13. What kind of mastery testing is specified in the program?

- Repeated items from student workbook
- Items similar to those in student workbook
- Items not similar to those in student workbook
- Standardized tests specified
- No mastery testing specified

5%
58%
5%
16%
9%

14. What kind of follow-up activities are specified for students who do poorly on the mastery tests?

- No follow-up exercises specified
- Lessons to be repeated
- Supplemental exercises, the same for any student below mastery level
- Supplemental exercises, different for different mastery test results
- Lessons to be repeated with specified criterion for those lessons

23%
13%
34%
21%
4%

Section III

1. About how many lessons in the level you teach focus on this skill?

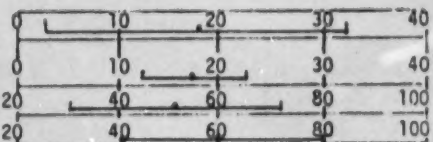
2. About how many minutes long are lessons or parts of lessons that focus on the main idea?

3. What percentage of the students knew main idea before the lessons?

4. What percentage of the students master main idea?

5. Here are some typical types of main idea exercises. Put a P by those that are in the program you use. Put a U by those that you think are particularly useful in teaching main idea. You may mark both P and U for any items

- Students underline main idea sentence when it is topic sentence.
- Students decide if first or last sentence tells the main idea.
- Students indicate which of three sentences tell the main idea.
- Students decide if paragraph contains a main idea sentence and, if so, they underline it.
- Students write their own main idea sentence.
- Teacher goes through paragraph one sentence at a time and asks if that sentence tells the main idea.
- Teacher asks students to tell the main idea and then has students try to find a main idea sentence in the paragraph.



P	U	UP

6. Indicate the percent of students that have these problems:

- confuse main idea with title
- think that main idea is first sentence
- think main idea must be a sentence that comes from the text
- cannot pick out the correct main idea sentence if it does not contain any phrase from the passage
- cannot generate main idea sentence
- other

30%
42%
46%
42%
36%
24%

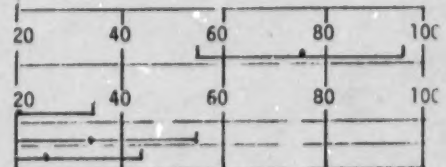
7. How closely do you follow the specified directions in your program for teaching main idea?

8. Indicate how you generally change your program's specific activities for teaching main idea. (Check all that apply.)

- I skip the following percentage of specified activities:
- I add the following percentage of activities:
- I change the following percentage of specified activities:

9. If you change specified activities, how do you change them? (Check all that apply)

- Provide more directed teaching
- Provide more review
- Do more testing
- Give more examples
- Give harder examples
- Give easier examples
- Spend more time on teaching the main idea



71%
55%
7%
79%
25%
32%
59%

BEST COPY AVAILABLE

10. How often is the topic of the main idea reviewed? 2 days 5 days 20 days 40 days
11. How often do you think main idea should be reviewed? 2 days 5 days 20 days 40 days

Section IV

1. Please rate the teaching difficulty of the following reading comprehension topics.

	easy		difficul
main idea			
cause/effect			
implied details			
vocabulary			
outlining			
reference skills			
reading for purpose			
following directions			
context clues			
drawing conclusions			
deductions			
determining mood			
point of view			
fact versus opinion			
determining plot			
analogies			
sequence of events			
predicting outcomes			
	agree		disagree
2. "... the school experience must help [students] develop positive attitudes toward themselves... This in turn provides a base on which the programs' other [academic] projects can be built."			
3. "... learning is likely to be more effective if it grows out of what interests the learner..."			
4. "... schools bring little influence to bear on a child's achievement that is independent of [his/her] background and social context..."			

for vocabulary). Interestingly, the average teacher rates outlining, implied details, and deductions as the easiest-to-teach skills (Section IV; 1), perhaps because the teacher's program does not deal with these skills in a way that would suggest their difficulty.

The teacher believes that the school has the potential to change the child's achievement, that learning should stem from learner interest, and that a positive attitude is the foundation for academic learning.

GRADE LEVEL DIFFERENCES OF QUESTIONNAIRE RESPONDENTS

Of the 65 items on the questionnaire that dealt with program, teaching, or student outcomes, nine show grade-level differences. Four have to do with general or specific teaching behavior, three relate to student outcomes, and two relate to program and topics taught in the program.

Below is a list of responses made by 4th grade teachers that are significantly different from those of 4th and 6th grade teachers (significant at the .05 level).

Fourth-grade Teachers Report

1. Spending more time on reading instruction.
2. Correcting a lower percentage of student errors.
3. Less frequent use of the correction procedure of telling the answer and repeating the question.
4. More frequent use of prompts when "students need help."
5. A lower percentage of students who knew main idea before the topic was introduced in the program.
6. A higher percentage of students who cannot identify the main idea if it does not contain words from the passage.
7. A higher percentage of students who exhibit other types of problems learning main idea.
8. Less frequent program review of main idea.
9. Cause/effect, reference skills, and fact-opinion are relatively harder to teach; determining plot is relatively easier to teach.

Some responses would be expected, such as the judgment that a lower percentage of fourth graders already know main idea, and that these students might have more problems with the various passage types for main idea. Some differences reflect that the teachers tend to teach the students in a "baby" manner, by prompting them more, correcting them less often, and correcting them less articulately.

These behaviors suggest that fourth grade teachers may be more naive about instruction than higher-grade teachers. They probably aren't provided with information about the outcomes of their instruction and therefore are less likely to appreciate either the necessity of some behaviors or the difficulty of some skills. The programs that are used may account for the relative ordering of how difficult topics are, perhaps because the treatment of some topics (such as cause/effect) in the fourth grade series is fairly simple and doesn't require the skill called for in later levels of the program.

In all, the differences by grade level underscore a repeated theme. There is no great difference between fourth grade and fifth grade or sixth grade with respect to what is taught, how it is taught, or what is mastered by the students.

COMPARISON OF QUESTIONNAIRE RESPONDENTS AND OBSERVED TEACHERS

After all data had been collected from the 17 sample teachers, they were asked to complete the same questionnaire that was mailed to 3085 teachers. Because of the late date only eight teachers in the study returned questionnaires.

Table 22 lists questionnaire responses, sample teacher responses, and comparison data for teachers in the study. Table 22 covers all but four questionnaire items. The comparable data column lists: (a) reported responses (from the first and second audio interviews, A1 and A2), and (b) actual data (based on the first and second video-tapings, V1 and V2; the program analysis (P1); student tests (S2) and the lesson-analysis of the video-taping (L2)).

Section I in Table 22 compares basic demographic and classroom information. Only on years of teaching experience is there a sizable difference.

Sections II and III of the questionnaire cover teacher behavior (as it applies to teaching any topic and to the topic of main idea). For these sections, there are four types of relationships between the groups.

1. Both groups achieve same scores, inaccurate answers.
2. Both groups achieve same scores but no accuracy data is available.
3. Groups achieve different scores but give inaccurate answers.
4. Groups achieve different scores, there is no accuracy data.

Section IV deals with the difficulty of teaching various topics and agreement with various quotes on education.

1. Same scores, inaccurate answers. On the following questions, there was little difference between responses of the groups. All responses were grossly inadequate. The mean response difference from the observed data score for questionnaire respondents and teachers, respectively, is indicated in parentheses.

TABLE 22 - COMPARISON OF QUESTIONNAIRE RESPONDENTS AND TEACHERS

SECTION I

	Questionnaire Respondents N=493	Observed Teachers N=8	Observed Teaching N=17	Other Compariso
3. How many students do you have for reading?	26	28		
4. How many years teaching experience do you have?	5.5	4.0		7.1 A2
5. How would you describe your reading class?	average	average		low av: A2
6. How would you describe the teaching instruction in your class?	sm. group	sm. group		sm. group n2
7. How many hours per week do the students spend in reading instruction	5.8	6.6		
8. How many hours per week do you spend actually teaching reading?	5.5	6.2		

SECTION II

1. In oral teacher directed activities how closely do you follow the procedures that are specified in the teacher's guide?	74%	74%	31% on main idea and cause/effect lessons	V2
2. If a good student makes a mistake on an orally-presented reading activity, what percentage of the time will at least one other student repeat the same mistake?	22%	31%		50% A2
3. If a low student makes a mistake, what percentage of the time will at least one other student repeat the same mistake?	18%	11%		
4. During an average lesson, what percentage of the student responses will be incorrect?	18%	11%	27%	V 1&2
5. As a generally acceptable guide, what percentage of the students should be able to master any particular skill in this reading program? ..	81%	84%		86% A2
6. What percentage of the mistakes made during oral presentations do you correct?	68%	88%	37%	V 1&2
7. Indicate the percentage of time you use each of these corrections:				
• Call on another student	51%	40%		37% acc A2
• Tell the student the answer is wrong and repeat the question	40%	47%		V 1&2
• Tell the answer and then repeat the question	19%	47%	32% acc	
• Tell the answer, repeat the question, then repeat the question to the whole group	22%	57%	50% acc (teach ques)	
• Ignore the mistake and permit the students to discover a natural consequence of their error	25%	15%	44% acc (resp)	
8. Most reading programs make general comments about praising and reinforcing students for correct responses. What percentage of the student responses do you reinforce in some way?	72%	86%	47%	78% A2
9. What percentage of the time do you call on individual students to answer questions (rather than the whole group or class)?	70%	76%		85% V 1&2
10. In what percentage of the lessons do worksheet exercises for a topic appear the same day as teacher activities for that topic?	66%	93%	12%	68% A1
11. There are places in the program where it says to help or guide the students. In what situations is this appropriate? (Check all that apply)				
Students have just learned the skill	79%	100%		54% acc A2
Students rarely use the skill and tend to forget it	55%	62%	52% acc (teach ques)	
Students need the help or they will never master the skill	59%	88%	49% acc (resp)	
Students have never really been taught the skill	54%	88%		
Students must answer an open-ended question	35%	25%		
12. How would you describe the teacher directions of your program?	53% spec. instruc.	30% spec. instruc.		

SECTION III

1. About how many lessons in the level you teach focus on this skill?			90% correct within 1/2 = A1
2. About how many minutes long are lessons or parts of lessons that focus on the main idea?	18%	8%	
	17%	18%	

BEST COPY AVAILABLE

TABLE 22 (cont.)

	Questionnaire Respondents N=493	Observed Teachers N=8	Observed Teaching N=17	Other Compar- isons
3. What percentage of the students knew main idea before the lessons? ..	32%	32%		44% A1
4. What percentage of the students master main idea?	60%	60%	10%	56% A1
6. Indicate the percent of students that have these problems:				
• confuse main idea with title	30%	36%		
• think that main idea is first sentence	42%	45%		
• think main idea must be a sentence that comes from the text	46%	54%		
• cannot pick out the correct main idea sentence if it does not contain any phrase from the passage	42%	43%		
• cannot generate main idea sentence	36%	44%		
7. How closely do you follow the specified directions in your program for teaching main idea?	76%	58%	31%	57% A1
8. Indicate how you generally change your program's specific activities for teaching main idea. (Check all that apply.)				
• I skip the following percentage of specified activities:	20%	40%	20% proc. var.	V2
• I add the following percentage of activities:	34%	52%	51% add. ques.	V2
• I change the following percentage of specified activities:	26%	50%	28% proc. var.	V2
9. If you change specified activities, how do you change them? (Check all that apply)				
Provide more directed teaching	71%	89%		
Provide more review	55%	75%		
Do more testing	7%	0%		
Give more examples	79%	89%		
Give harder examples	25%	38%		
Give easier examples	32%	25%		
Spend more time on teaching the main idea	59%	50%		
10. How often is the topic of the main idea reviewed?	11 days	20 days	62	P1
11. How often do you think main idea should be reviewed?	5 days	4 days		

Section IV

1. Please rate the teaching difficulty of the following reading comprehension topics.		indicates very difficult to teach.	
Questionnaire Responses	Observed Teachers	Quest. Response	Observed Teachers
main idea	2.2	2.2	2.4
cause/effect	2.2	1.9	1.9
implied details	1.9	2.4	2.3
vocabulary	3.3	2.4	3.0
outlining	1.8	3.0	3.1
reference skills	2.9	2.5	2.0
reading for purpose	2.9	2.1	2.9
following directions	2.6	2.6	2.6
context clues	2.8	2.7	2.9
drawing conclusions			
deductions			
determining mood			
point of view			
fact versus opinion			
determining plot			
analogies			
sequence of events			
predicting outcomes			

- "... the school experience must help [students] develop positive attitudes toward themselves... This in turn provides a base on which the programs' other [academic] projects can be built."
- "... learning is likely to be more effective if it grows out of what interests the learner..."
- "... schools bring little influence to bear on a child's achievement that is independent of [his/her] background and social context..."

1.4	1.0
1.6	1.5
3.1	3.3

	<u>Questionnaire respondents</u>	<u>Teachers</u>
II, 1. How closely do you follow specified teaching procedures?	estimate 35% above observed teaching	estimate 43% above observed teaching
II, 4. What percent of the responses will be incorrect?	estimate 9% below observed teaching	estimate 16% below observed teaching
II, 8. What percent of responses do you reinforce?	estimate 26% above observed teaching	estimate 40% above observed teaching
II, 9. What percent of your questions are answered by individual students?	estimate 15% below observed teaching	estimate 9% below observed teaching
III, 4. What percent of the students master main idea?	estimate 50% above observed teaching	estimate 50% above observed teaching
III, 10. How often is main idea reviewed? (46 days variance from actual frequency)	estimate 51% above observed teaching	estimate 42% above observed teaching

2. Same scores, no accuracy. On the following questions, there was little difference between the scores, but there was no comparable data to check the accuracy.

- II, 5. What percent should master any skills?
- II, 2. If a good student errs, how many will repeat the error?
- II, 3. If a poor student errs, how many will repeat the error?
- III, 2. How long are main idea lessons?
- III, 3. What percent of the students know main idea before the lesson?
- III, 6. What percent of the students have these different problems with main idea exercises?
- III, 11. How often should main idea be reviewed?

3. Different scores, inaccurate answers. On the following sections, there is a sizeable difference between the responses of the observed teachers and

those of the questionnaire respondents. However, both average responses are inaccurate.

Table 23 compares the responses of observed teachers and the questionnaire respondents with the actual teaching behavior of the observed teachers.

		Table 23			
Item		Questionnaire Respondents	Deviation From Actual Behavior	Observed Teachers	Deviation From Actual Behavior
II, 6	What percentage of errors do you correct?	68%	+31%	85%	+53%
II, 10	What percentage of the time is teacher and student work for topic on same lesson?	66%	+60%	93%	+81%
III, 7	How closely do you follow specified directions for teaching main idea?	76%	+35%	58%	+27%

As Table 23 shows, the percentages are overestimated by both groups. The sample of observed teachers overestimates more than the questionnaire group on the first two items while the questionnaire group overestimates more on the last question. Note, however, that the trend is in the same direction, that of over-estimation.

The following items show the difference between the two groups. Direct data on accuracy is not available.

- II, 12 How would you describe the directions of your program?
(Teachers indicated instructions were 23% less specific.)
- II, 7 What percentage of the time do you use these corrections? (44% accurate/12% too high; 50% accurate/18% too high)
- III, 1 How many lessons teach main idea? (Observed teachers indicated half as many as respondents.)
- III, 8 How do you generally change program-specified activities?
(Observed teachers indicated greater percentages of changes.)
- III, 9 How do you specifically change those activities? (Observed teachers generally indicated higher percentage of changes.)

Item II, 7 deals with corrections. There are observation data on corrections. By using the fact that observed teachers corrected only 32% of the mistakes, we could judge the overall responses to II, 7 as an overestimation of the use of procedures. The overestimation could be expressed as 12% overestimation for the questionnaire group and 18% overestimation for the observed group. These combined-mean overestimations are quite close to each other, even though the two groups vary substantially on three of the five items in II, 7.

Section IV asked teachers to judge how difficult it is to teach certain topics and how teachers felt about three different quotes on education.

Differences in judged difficulty appeared between the groups on main idea, vocabulary, following directions, point of view and analogies. On the remaining 13 topics, there was little if any difference between the respondents and the sample teachers. There were very slight differences in responses to the quotes, but both groups strongly agreed with the first two quotes and did not endorse the last quote.

Implications

The sample of questionnaires filled out by teachers who had been observed is unfortunately small. Furthermore, the sequence of events is not conducive to achieving uncontaminated responses from the observed teachers. If the questionnaire had been presented, first, before the audio interviews (or perhaps in place of parts) and before teaching, we would be able to make a much stronger case about whether the observed teachers are typical of the larger population. As it is, we can make the following suggestions.

1. Those observed teachers who did respond to the questionnaire are typical of the other observed teachers with respect to: (a) grade level, (b) teaching performance, (c) student outcomes, and (d) verbal responses.
2. On 46 items of the questionnaire, the sample of observed teachers produce the same type of responses as the larger group of questionnaire respondents.

3. On three items, there are sizable differences in the type of response; however, the trend is in the same direction. (If one group underestimates, the other group underestimates.)
4. On 16 items, there are differences that are not accounted for by the present study. For instance, the sample of observed teachers rate main idea much easier to teach than the larger group of questionnaire respondents. Such distortion may result from the "professionalism" and focusing of attention that results from discussing these topics and teaching them in situations that made it obvious that the teacher's performance was being observed. Deviations of this sort are possible on any of the issues that have been discussed, because the discussion serves as a form of contamination, a treatment that is not received by the other respondents. In other words, the responses of the sample teachers may be highly similar to the other group's only because it had been distorted by the study. The striking similarity of the two groups responses on most items, however, militates somewhat against this interpretation.

If we assume that the teachers who were observed comprise a reasonably typical sample of teachers (an assumption that the investigators espouse), we may draw the following conclusions about the typical teacher in grades 4 through 6.

The teacher agrees with statements that refer to specific instructional problems, such as the lack of feedback procedures.

These endorsements of the problems are generally "accurate" which means that the teacher actually does know the suggested problem.

The teacher becomes vague when dealing with statements about solving the problem.

This vagueness generally predicts lack of understanding of how to solve the problem.

The general implication seems to be that the teacher can be relied on to provide honest responses if problems are presented as specific problems--not as jargon or catch phrases. The teachers know, for example, that students have trouble confusing main idea and the first sentence of the selection. So long as we deal with specifics, we will get honest answers. Only when we deal with vague topics or those that are beyond the teacher's ken do we start getting unreliable responses. This fact is supported by the inter-correlations of items in the questionnaire. One of the few high correlations was between the judged difficulty of teaching "determining mood" and that of "point of view." Teachers tend to respond in the same way to these items, suggesting that perhaps the teachers don't understand how they are different. For those items with which teachers most probably have experience, however, the correlations are very low, implying that the teachers provide independent judgments for each item (which means that their responses are probably reliable judgments, if not completely accurate).